UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Development of an Agriculture Drought Risk model for the Iberian Peninsula

" Documento Definitivo"

Doutoramento em Ciências Geofísicas e da Geoinformação
Especialidade de Meteorologia

Andreia Filipa Silva Ribeiro

Tese orientada por:
Ana Cristina Machado Russo
Célia Marina Pedroso Gouveia

Documento especialmente elaborado para a obtenção do grau de doutor

2020

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Development of an Agriculture Drought Risk model for the Iberian Peninsula

Doutoramento em Ciências Geofísicas e da Geoinformação

Especialidade de Meteorologia

Andreia Filipa Silva Ribeiro

Tese orientada por:

Ana Cristina Machado Russo

Célia Marina Pedroso Gouveia

Júri:

Presidente:

- Doutor João Manuel de Almeida Serra, Professor Catedrático e Presidente do Departamento de Engenharia Geográfica, Geofísica e Energia da Faculdade de Ciências da Universidade de Lisboa.

Vogais:

- Doutor Sérgio Vicente-Serrano, Investigador Científico, Instituto Pirenaico de Ecología (Espanha);
- Doutora Susana Alexandra Barbosa, Investigadora Sénior, INESC-TEC, Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência;
- Doutor Carlos do Carmo de Portugal e Castro da Câmara, Professor Associado, Faculdade de Ciências da Universidade de Lisboa;
- Doutor Ricardo Machado Trigo, Professor Associado, Faculdade de Ciências da Universidade de Lisboa;
- Doutora Ana Cristina Machado Russo, Investigadora Júnior, Faculdade de Ciências da Universidade de Lisboa (orientadora).

Documento especialmente elaborado para a obtenção do grau de doutor

Fundação para a Ciência e Tecnologia (Bolsa PD/BD/114481/2016)

2020

All my time and efforts to this venture is lovingly dedicated to André Ricardo Amaral.

Acknowledgments

It is with great pleasure and gratitude that I acknowledge the role of those who have supported me, each in its own way, along this scientific and personal pilgrimage.

I owe my deep and wholehearted gratitude to my supervisors Ana Russo and Célia Gouveia. These four years were some of the best of my life and you played a big role on my happiness and both professional and personal accomplishment. I always felt we were on the same page and you always believed in me much more than myself. I was blessed to be constantly guided, instructed, advised, supported, encouraged, comforted, by you. Besides your expertise, know-how, dynamism and vision, I'm also grateful for your generosity and kindness in both weakest and best moments. Thanks to your dedication this journey was constructive and fruitful and most of all, exciting and enjoyable. I also gratefully thank you for including me in the project IMPECAF (PTDC/CTA-CLI/28902/2017) which co-funded this PhD research together with the grant PD/BD/114481/2016 from the Portuguese Science Foundation (FCT).

I became a member of FCUL in 2007, and posteriorly a member of IDL in 2012, that besides being my workplace for the last couple of years, allowed me to grow as a researcher and as a person, surrounding me with people and relationships that I will carry with me into the future. First, the greatest thanks to my friends and role model climate researchers, Sofia Ermida and Patrícia Páscoa, for all the friendship, your valuable insights and great discussions. Also, this way would not have been so great without those who started their PhD with me in IDL: Riccardo Hénin, Joana Carvalho, Joana Ferreira, Filipa Luz, Alessandro Righetti, Dorota Szalaj and João Dias. The lunches, coffee breaks and "team-building" moments are no less important, so many thanks to Silvia Nunes, Pedro Sousa and to the Earthsystems and Sustainable Energy Systems doctoral program's circles. Finally, many thanks to all the IDL researchers and professors that closely followed my steps during this thesis, and in many ways have supporting and guiding me, namely: Carlos Pires, Carlos da Câmara, Ricardo Trigo, Alexandre Ramos and Margarida Liberato.

Networking was a very important part of this PhD. I would like to acknowledge the members of the COST Action CA17109 (DAMOCLES) and the team of the Climate and Environmental Physics in the University of Bern. Thanks for everything I have experienced and learned with you, and for welcoming in Bern during the Short-term Scientific Mission. A special note of

gratitude to Jakob Zscheischler for the guidance and inspiration, “Statistic is the answer, let it be, let it be”. Also, financially, I would like to acknowledge the CA17109 and the University of Bern for co-funding the research conducted in Switzerland.

My family and oldest friends have always been my fort, my shelter, my safe port, during my whole life. A very special one left this physical world during the very last steps of the writing of this manuscript, to be a shining star in the sky and in my heart, my grandmother Emília. Thank you for all the seconds of my life with you, and for your words of love and wisdom that will always be a precious treasure to me. A very special thank you to my parents and my sweet sister (the true joy of my life) who have always taken care of me, supported and encouraged me, with such love, to follow my passions. Particularly because I know that is difficult to understand the research life and how this job and the funding works, and still you gave me unconditional support. I extend this gratitude to all my lovely and caring family, thank you very much for the prayers and true love and for forgiving me my absence many times. Also the deepest thanks to my dearest childhood and adulthood friends, you are also my family.

André, this thesis is dedicated to you. Thank you for believing in me, and for remember me who I am and who I want to be, all the time. You are one of the greatest minds that I know, and you always helped me to keep things in perspective. Also, thank you for proofreading my texts several times, for the endless emotional support and for telling me “the world needs you” when I’m away. You make me a better person and researcher, thank you.

Last of all, in gratitude to God my heart sings:
Kanimambo Hosi Yanga,
Kanimambo Tatana,
Kanimambo Murisi wanga,
Kanimambo Xikwembu Xanga!

Abstract

Among climate extremes, droughts are a major source of risk to agriculture and food security, which are expected to be increasingly affected considering the tendency towards a warmer climate. Within the context of climate change, the Iberian Peninsula (IP) is one of the regions recurrently highlighted as one of the areas expected to be particularly affected by drought episodes, due to the strong variations in the precipitation regime that make the region prone to drought events. In this way, this dissertation aimed the development of an agricultural drought risk model to contribute to more resilient systems in the IP.

The skills of several drought indicators (SPEI, VCI, TCI and VHI) in predicting wheat and barley yields were firstly assessed based on neural networks and multiple linear regression models. Afterwards, copula-based models were designed to assess the joint probability of crop yields and droughts for a probabilistic risk assessment. The agricultural drought risk was then defined as the conditional probability of crop-loss under drought conditions and mapped at the province level of the IP. Ultimately, the additional risk associated with the occurrence of extreme temperatures during droughts was evaluated to characterize how the interaction between dry and hot conditions may exacerbate the impacts of the individual hazards in agriculture.

The results showed the good performance of drought indicators in predicting the occurrence of crop failures. In general, barley exhibits greater agricultural drought risk in comparison to wheat. Overall, the risk of crop-loss increases with the severity of drought conditions, and drought-related risks increase with the interaction with extreme temperatures. Although compound dry and hot conditions lead to the larger damages in crop yield than the individual drought- or heat-stress, drought is still the dominant factor. From an operational point of view, this research intends contributing to the agricultural decision-making.

Keywords: Drought-related yield losses; Iberian Peninsula; Wheat and barley; Multivariate statistical models; Extreme events

Resumo

Entre os extremos climáticos, as secas destacam-se atualmente como uma fonte de risco relevante para os sistemas agrícolas e de segurança alimentar, e vão continuar a sê-lo no futuro, em consequência do aquecimento global. Particularmente, a Península Ibérica (PI) é uma região propensa à ocorrência de secas com impactos na agricultura em consequência da grande variabilidade do regime de precipitação e pela predominância de práticas agrícolas de sequeiro em relação ao regadio. Neste sentido, a avaliação do risco na agricultura associado a eventos de seca é de elevada importância nesta região, já que poderá constituir uma ferramenta poderosa para o estabelecimento de sistemas agrícolas mais resilientes.

O principal objetivo desta tese foi contribuir para o desenvolvimento de um modelo de avaliação de risco em culturas cerealíferas de sequeiro associado a episódios de seca na PI. A tese está estruturada em torno de quatro artigos científicos que tinham o objetivo de analisar em detalhe a resposta da produtividade agrícola às condições de seca e desenvolver modelos estatísticos para estimar os riscos de perdas nas colheitas associados a eventos de seca. A resposta a estas questões constitui uma contribuição significativa para um conhecimento mais aprofundado da quantificação do risco na agricultura relacionados com a seca, pelo recurso a metodologias estatísticas inovadoras capazes de capturar o comportamento extremo deste fenómeno, tendo em conta os impactos causados.

No primeiro artigo foi avaliada a capacidade de previsão da produtividade de colheitas de sequeiro (trigo e cevada) através de vários indicadores de seca, nomeadamente, o SPEI (Standardized Precipitation Evaporation Index), o VCI (Vegetation Condition Index), o TCI (Temperature Condition Index) e o VHI (Vegetation Health Index). A influência das condições de seca nas culturas de cereais na PI foi avaliada, considerando dois clusters de províncias dominadas por agricultura de sequeiro. Esta análise foi conseguida com base em análise de correlação, técnicas de regressão linear e redes neuronais artificiais, para várias escalas temporais e diferentes momentos do ciclo vegetativo. Os resultados da análise de correlação com base nos indicadores de seca, mostraram que estes eventos têm maior influência durante as etapas do ciclo vegetativo em que a vegetação é fotossinteticamente mais ativa (primavera e início do verão). Os modelos de regressão linear e de redes neuronais revelaram que o desempenho das diferentes metodologias é bastante semelhante. De um modo geral, as perdas

de produtividade relevantes foram previstas a partir dos modelos estimados (mais de 69% de taxa de acertos).

No segundo artigo foram utilizadas funções cópulas bivariadas para caracterizar a estrutura de dependência entre a produtividade e as condições de seca através da sua distribuição multivariada. A aplicação de cópulas elípticas e arquimedianas para estimar distribuições de probabilidade conjunta é recente em estudos agrometeorológicos. Os resultados indicaram que as cópulas arquimedianas são as que melhor caracterizam o comportamento conjunto entre produtividade de cereais e os eventos de seca, sugerindo a existência de uma dependência nos valores extremos destas variáveis. O uso de cópulas permitiu o cálculo de probabilidades condicionadas de perdas de produtividade em situação de seca e a comparação com a probabilidade de perda de produtividade sem constrangimento de seca. Os resultados mostraram que o risco de perda nas colheitas de trigo e cevada pode ser subestimado quando a probabilidade de perda não é condicionada à situação de seca. Os valores de probabilidade condicionada de perda de produtividade sugeriram que o risco aumenta quando as condições evoluem de normais ou húmidas para condições de seca moderada ou severa. Ao nível dos conjuntos de províncias, os resultados mostraram que a ocorrência de seca acarreta um maior risco para a cevada na região a norte da PI. Em contraste, na região a sul da PI o risco é maior para o trigo, sugerindo que a escolha dos cereais para cultivo na região a norte (sul) deveria dar preferência ao trigo (cevada), visando a redução dos riscos de perdas associados a secas.

No terceiro artigo, foi definida uma medida de risco interpretável para apoio à decisão, onde o risco na agricultura associado a eventos de seca corresponde à probabilidade condicional de perda de produtividade induzida por episódios de seca. Os resultados dos trabalhos anteriores com base nas funções cópulas foram aplicados a todas as províncias da PI e foram construídos mapas de risco em termos de perdas de produtividade de trigo e cevada para diferentes condições de severidade da seca. Os resultados indicaram que na maior parte das províncias o risco aumenta com a severidade da seca, e a cevada apresenta maiores valores de risco em mais províncias do que o trigo. De um modo geral, a estrutura de dependência entre produtividade e seca exibe uma dependência nos valores extremos. Enquanto o índice TCI está mais associado a funções cópulas com dependência nos valores extremos mais elevados (cópulas Gumbel), os índices VCI e SPEI estão principalmente associados a funções cópulas com maior dependência nos valores extremos mais baixos (cópulas Clayton). Neste estudo, as províncias da Galiza destacaram-se como províncias com um risco bastante baixo no caso dos dois cereais, em

comparação com as restantes províncias. Salamanca revelou ser a província com maior risco na agricultura associado a eventos de seca.

Por fim, no quarto artigo, o risco adicional associado à ocorrência de temperaturas extremas durante as secas foi avaliado com o intuito de caracterizar em que medida o papel desempenhado pela interação entre extremos climáticos pode exacerbar os riscos individuais nos sistemas agrícolas. Neste estudo foi estimada a dependência tripla entre temperatura máxima, precipitação e produtividade de cereais com base em cópulas arquimedianas a três dimensões. A partir das distribuições conjuntas trivariadas foram estimadas probabilidades condicionais de perdas de produtividade de trigo e cevada induzidas pela interação entre seca e calor e comparadas com o risco individual de cada um dos extremos climáticos. Os resultados sugeriram que o risco de perdas agrícolas aumenta com o aumento da severidade da seca e calor simultâneos e que o risco induzido unicamente por um dos extremos climáticos (seca ou calor) é amplificado pela interação dos mesmos. Além disso, os resultados indicaram que a ocorrência de seca tem um papel predominante no risco induzido pela interação entre seca e calor, sugerindo que a seca não precisa de ser tão extrema quanto os extremos de temperatura para causar danos equivalentes na produtividade. A cevada revelou ser o cereal que apresenta maior risco de perda de produtividade associado à interação dos extremos de seca e temperatura, à semelhança dos estudos anteriores focados apenas na seca.

Em anexo encontra-se um estudo paralelo ao trabalho desta tese, focado na ocorrência de extremos de calor induzidos pelas condições de seca anteriores. Os resultados mostraram que regimes de seca nos meses anteriores aumentam a probabilidade de ocorrência de temperaturas extremas no verão e que a ocorrência de ondas de calor pode ser antecipada pelas condições de seca antecedentes.

Globalmente, esta tese propõe o recurso a uma avaliação combinada do risco de seca e dos impactos associados, capturando o caráter multivariado e extremo do risco de seca na agricultura. De um ponto de vista operacional, pretende-se que estes estudos contribuam para os processos de tomada de decisão na agricultura, evidenciando as vantagens da escolha diferenciada das culturas bem como do recurso a seguros agrícolas que visem colmatar possíveis consequências socioeconómicas.

Palavras-chave: Impacto das secas na agricultura; Península Ibérica; Trigo e cevada; Análise estatística multivariada; Eventos extremos.

Contents

List of Acronyms and Abbreviations

IP	Iberian Peninsula
SPEI	Standardized Precipitation Evapotranspiration Index
SPI	Standardized Precipitation Index
NDVI	Normalized Difference Vegetation Index
VCI	Vegetation Condition Index
TCI	Temperature Condition Index
BT	Brightness Temperature
VHI	Vegetation Health Index
PDSI	Palmer Drought Severity Index
sc-PDSI	Self-calibrated Palmer Drought Severity Index
SSI	Standardized Soil-moisture Index
FAO	Food and Agriculture Organization
IPCC	Intergovernmental Panel on Climate Change
MLR	Multiple Linear Regression
ANN	Artificial Neural Networks
HR	Hit Rate
FB	Frequency Bias
SR	Success Ratio
FAR	False Alarm Rate
GDP	Gross Domestic Product
NAO	North Atlantic Oscillation
INE	Instituto Nacional de Estatística
CORINE	Coordination of information on the environment
CLC	CORINE Land Cover
NOAA	National Oceanic and Atmospheric Administration
NOAA-AVHRR	NOAA - Advanced Very High-Resolution Radiometer
ftp	File transfer protocol
CRU	Climate Research Unit
PET	Potential Evapotranspiration
RMSE	Root Mean Squared Error

CERES	Crop Environment Resource Synthesis
AquaCrop	Crop growth model developed by FAO
RCP8.5	Representative Concentration Pathway (8.5 Wm^{-2} radiative forcing)
CDF	Cumulative Distribution Function
MLE	Maximum Likelihood Estimate
IFM	Inference Functions for Margins
CML	Canonical Maximum Likelihood
AIC	Akaike's Information Criteria
PNE	Probability of Non-Exceedance
CPNE	Conditional Probability of Non-Exceedance
ECDF	Empirical Cumulative Distribution Function
PDF	Probability Distribution Function
EU	European Union
AC	Archimedean Copulas
HAC	Hierarchical Archimedean Copulas
NAC	Nested Archimedean Copulas
LOESS	Locally Estimated Scatterplot Smoothing
MAM	March, April, May
QQ	Quantile-Quantile
USA	United States of America
US	United States
UK	United Kingdom
IM	Instituto de Meteorologia
AEMET	Agência Estatal de Meteorologia
NHD	Number of Hot Days
NW	Northwestern
NE	Northeastern
SW	Southwestern
SE	Southeastern
C	Central
W	Western
NASA	National Aeronautics and Space Administration

SDHI	Standardized Dry and Hot Index
ECAD-EOBS	European Climate Assessment and Dataset E-OBS
PCA	Principal Component Analysis
BIC	Bayesian Information Criterion
RCM	Regional Climate Model
GCM	Global Climate Model
IDL	Instituto Dom Luiz
STSM	Short-term Scientific Mission
COST	European Cooperation in Science and Technology

List of Symbols

$NDVI_{min}$	Historical Minimum NDVI for the considered week
$NDVI_{max}$	Historical Maximum NDVI for the considered week
BT_{max}	Historical Minimum BT for the considered week
BT_{min}	Historical Maximum BT for the considered week
P	Precipitation
PET	Potential evapotranspiration
D	Water deficit between P and PET
R	Pearson correlation coefficient
R^2_{adj}	Adjusted coefficient of determination
$R^2_{adj\_no\_cv}$	Adjusted coefficient of determination without cross-validation mode
SS_{RMSE}	Skill Score based on the RMSE
std	Standard deviation
p	Number of predictors
X, Y	Observation vectors
x, y	Pairs of observations
F_X	CDF of X
F_Y	CDF of Y
F_{XY}	Joint CDF of X and Y
τ	Kendall's correlation
ρ	Pearson's correlation
C	Copula function
θ	Copula dependence parameter
df	Degrees of freedom
u, v	Marginal distributions
u_{sim}, v_{sim}	Simulated marginal distributions
Φ	Standard Gaussian CDF
t_{df}	t–student CDF
∞	Infinity
ln	Natural logarithm

X_{sim}, Y_{sim}	Simulated X and Y
$-X_{std}$	Threshold of minus one standard deviation of crop yield anomalies
$PNE_{-X_{std}}$	Probability of non-exceedance -1 std of crop yield anomalies
Y_{th-dr}	Threshold of drought condition
$CPNE_{-X_{std}\|drought}$	Conditional probability of crop-loss under drought conditions
$CPNE_{-X_{std}\|non\_drought}$	Conditional probability of crop-loss under non-drought conditions
v_{th-dr}	Marginal probability of drought occurrence
u_{-std}	Marginal probability of crop-loss
$F_{X_{sim}}$	Probability distribution of X_{sim}
$F_{Y_{sim}}$	Probability distribution of Y_{sim}
y^*	Threshold of drought condition (Chapter 4) Threshold of crop-loss (Chapter 5)
X_{sim^*}	Simulated X under drought condition
$F_{X_{sim^*}}$	Conditional probability distribution of crop-loss under drought conditions
std	Standard deviation
ci	Confidence interval
Tmax	Maximum temperature
N	Sample size
P_{MAM}	Spring precipitation
$Tmax_{MAM}$	Spring maximum temperature
x_1^*	Threshold of dry conditions
x_2^*	Threshold of hot conditions
$F_{Y\|X_1}$	Probability distribution of yield under dry conditions
$F_{Y\|X_2}$	Probability distribution of yield under hot conditions
$F_{Y\|X_1,X_2}$	Probability distribution of yield under compound dry and hot conditions
φ	Generator function of Archimedean copulas
+dry	Moderate dry conditions

++dry	Severe dry conditions
+++dry	Extreme dry conditions
+hot	Moderate hot conditions
++hot	Severe hot conditions
+++hot	Extreme hot conditions
+dry+hot	Moderate compound dry and hot conditions
++dry++hot	Severe compound dry and hot conditions
+++dry+++hot	Extreme compound dry and hot conditions
α	Significance level
$v_{sim,dry}$	NHD samples under drought conditions
$v_{sim,wet}$	NHD samples under normal/wet conditions
$F_{v_{sim,dry}}$	CDF of $v_{sim,dry}$
$F_{v_{sim,wet}}$	CDF of $v_{sim,wet}$
λ_U	Upper tail dependence
q	Quantile
H_0	Null hypothesis
C_θ	Parametric copula
C_n	Empirical copula

List of Figures

List of Tables

Chapter 1

Introduction

1. Introduction

The current section, Chapter 1, presents a general introduction to droughts, its associated agricultural impacts in the Iberian Peninsula (IP) and to the conceptual principles on risk analysis. The following chapters are *ipsis verbis* versions of published (Chapters 2, 3 and 4) and under review (Chapter 5) works, and the body of its original versions is here preserved. A graphical abstract is provided in the beginning of each chapter to give the reader an overview of each section's workflow. Each one of the Chapters 2, 3, 4 and 5 include, after the corresponding graphical abstract, the respective introduction, data and methods section, followed by results, discussion and conclusions. This applies also to the contents of the appendix section (Appendix A). To some extent, repetition of concepts, datasets and methods is unavoidable and, therefore, Chapter 1 mainly intends to contextualize the research questions addressed in the succeeding chapters. A more detailed description of the research objectives and thesis structure follows in the last sections of this opening chapter.

1.1 General concepts

1.1.1 Droughts and agricultural impacts

During the last century, climate change contributed to the exacerbation of the vulnerability of various natural systems. Drought is an extreme event of the climatic system with environmental, social and economic impacts, which differs from region to region over a range of temporal scales (Wilhite and Glantz, 1985; Wilhite, 2000; Zargar et al., 2011). Despite the lack of a consensual and objective drought definition, in the most general one drought is commonly known to result from a deficiency of precipitation over an extended period of time (Wilhite and Glantz, 1985; Wilhite, 2000; Zargar et al., 2011). Nevertheless, the success of the anticipation of drought consequences and their mitigation is largely dependent on a continuous monitoring of drought conditions.

This monitoring can be accomplished through different approaches and datasets, integrating multiple information sources, namely in-situ or remote sensing based methods, numerical models outputs, analysis and reanalysis or the combinations of different datasets. While most traditional methods of drought monitoring based on in-situ observations are better suited to

local scales, the recent advances of remote sensing offer the main advantage of studying larger areas (Lakshmi, 2016). On the other hand, numerical models simulating the interactions of the atmosphere, ocean and surface, based on the integration of physical, chemical, and sometimes biological equations, allow the diagnosis and prognosis of climate dynamics (Kalnay, 2003; Warner and Thomas, 2011). In this way, numerical models offer the advantage of generating historical and future climate data, including short-medium range forecast and projections of climate change (Kalnay, 2003; Warner and Thomas, 2011). In addition, global gridded climatological datasets generated from observations and assimilated data, also provide complete records of weather data (Cornes et al., 2018; Harris et al., 2020).

There are several examples of variables from these multiple information sources, which are used to characterize drought conditions, such as: precipitation, temperature, evapotranspiration, soil moisture, streamflow, groundwater, solar radiation, among others (Figure 1.1). These variables are often used individually or combined in the form of drought indices (may also be called indicators), which consist in representing drought conditions by a number and constitute useful tools for comprehensive drought monitoring via e.g. threshold triggers (e.g. percentile approach).

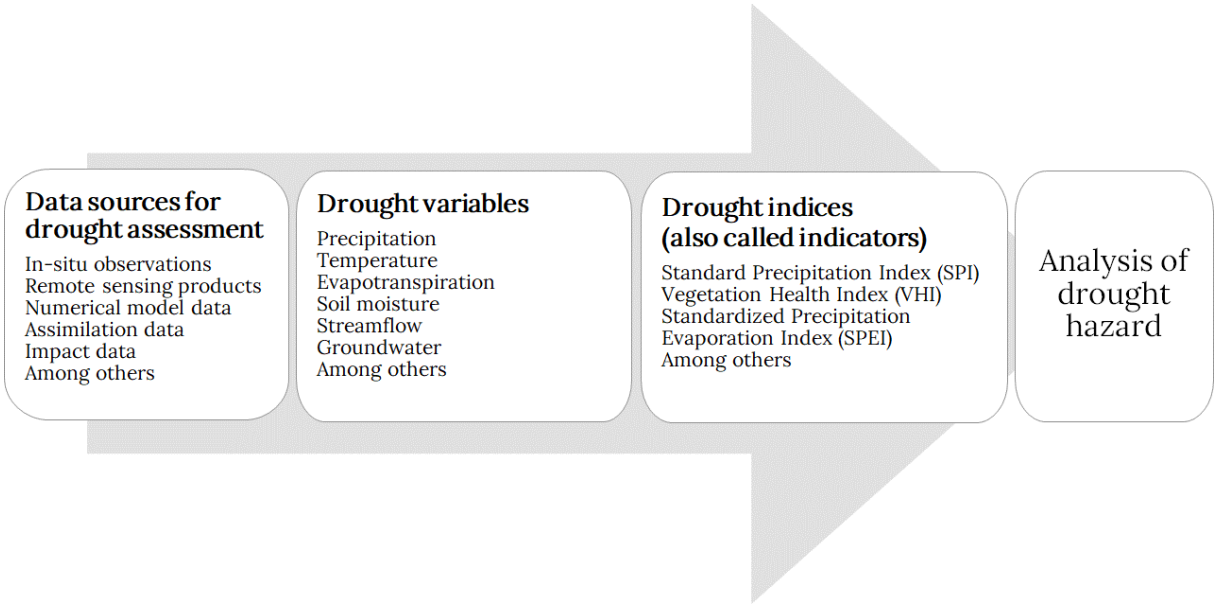


Figure 1.1 – General conceptional view of drought monitoring to carry drought hazard analysis.

A panoply of drought indices, with different characteristics, has been developed over the years including, inter alia, the Palmer Drought Severity Index (PDSI) (Palmer, 1965), the Standard Precipitation Index (SPI) (McKee et al., 1993), the self-calibrated PDSI (sc-PDSI) (Wells et al.,

2004) and the Standardized Precipitation Evaporation Index (SPEI) (Vicente-Serrano et al., 2010) (Table 1.1). While most indices try to characterize drought considering precipitation deficits or simple water balance deficits (e.g. PDSI, sc-PDSI), some do not have the ability of accounting for the multiscale characteristics of drought. In this sense, multiscalar drought indices have revealed to be extremely useful for drought assessment, due to their ability to evaluate the response time of different systems to drought events. In particular, the SPI has been widely used (Wu and Wilhite, 2004; Moreira et al., 2015; Gouveia et al., 2016a; Páscoa et al., 2017a). Although the traditional methods of drought monitoring rely on precipitation only as input (such as the SPI), a single variable may not be suitable for all regions and systems to characterize droughts. Hence a collection of multivariate drought indices has been developed (Figure 1.1 and Table 1.1), and the incorporation of impact data (e.g. water quality, crop yield loss) may also be viewed as a reliable indicator of drought conditions.

In order to overcome the use of an index exclusively based on precipitation, Vicente-Serrano et al. (2010) proposed the SPEI, which combines the influence of precipitation and evapotranspiration. Besides their multiscalar character, standardized indices like SPI and SPEI are particular recommended in drought analysis as they allow to perform comparisons between different regions and to provide a probabilistic interpretation (Guttman, 1998; Vicente-Serrano et al., 2010; Beguería et al., 2014). This is particularly true for the analysis of droughts in agricultural systems, which is one of the activities most directly influenced by climatic conditions, and consequently seriously affected by the occurrence of extreme events like droughts (Wilhite, 2005).

The identification of drought consequences in the agricultural land has different response times depending on the vegetation type, due to different levels of resistance and resilience to water stress (Wilhite, 2000; Vicente-Serrano et al., 2013). Typically, shorter timescales, as the timescale of 3-month, are useful for characterization of agricultural droughts, while longer time scales (e.g. 12-months) are more suitable when studying hydrological droughts (Vicente-Serrano et al., 2006). Moreover, the time-scales of 6 and 9 months in May and June have shown to be significant in the case of wheat yield for some IP provinces (Páscoa et al., 2017b) and thus, the assessment of the best time scales of drought indices must be performed for different moments of the vegetative cycle of the selected crops.

Table 1.1 - Examples of popular drought indices which stand out for their wide use. See Zargar et al. (2011) for a review of drought indices.

Drought Index	Brief description	Input data	Author
Palmer Drought Severity Index (PDSI)	Is a measure of soil moisture, derived from precipitation and air temperature. Shows a great importance in hydrology and agriculture because it considers the availability of soil water to vegetation.	Precipitation Temperature	Palmer (1965)
Standard Precipitation Index (SPI)	Is based on a probabilistic approach of the precipitation. Featuring long-term series of rainfall, this index allows assessing the probability of the precipitation being less than or equal to a certain threshold.	Precipitation	McKee et al. (1993)
Self-calibrated Palmer Drought Severity Index (sc-PDSI)	Developed with the aim of improving the performance of the PDSI reducing the excessive frequency of extreme events.	Precipitation Temperature	Wells et al. (2004)
Standardized Precipitation Evaporation Index (SPEI)	Similar to SPI and in addition takes into account the variability of evaporation caused by temperature fluctuations.	Precipitation Evapotranspiration	Vicente-Serrano et al. (2010)

Table 1.2 - Examples of drought-related remote sensing indices.

Drought-related remote sensing index	Brief description	Input data	Author
Normalized Difference Vegetation Index (NDVI)	Uses changes in vegetation cover as an indicator of droughts. NDVI uses infrared and near-infrared information to quantify vegetation stress, being sensitive to changes in liquid water content of vegetation canopies.	Visible red and near infrared bands	Multiple authors employ NDVI, e.g. Gouveia et al. (2009)
Vegetation Condition Index (VCI)	Mathematically expressed by NDVI relative to the minimum and maximum historical limits and further normalized relative to their amplitude interval; characterizes the moisture conditions of vegetation.	NDVI	Kogan (1995, 1990)
Temperature Condition Index (TCI)	Similar to VCI replacing the NDVI by the Brightness Temperature (BT); characterizes the thermal conditions of vegetation.	BT	Kogan (1995, 1990)
Vegetation Health Index (VHI)	Average between VCI and TCI in order to consider their combined effect on vegetation health.	VCI TCI	Kogan (1997)

In addition to hydro-meteorological drought indices like SPI and SPEI, the recent advances of remote sensing have strongly contributed to the monitoring of droughts in the agricultural sector (Kogan et al., 2015a; Bokusheva et al., 2016; Gouveia et al., 2017a). Among remote sensing indices for monitoring vegetation activity and drought conditions (Table 1.2), the Normalized Difference Vegetation Index (NDVI) is widely used as a measure of the greenness of the vegetation. Strong correlations between the NDVI and in both Portuguese and Spanish wheat yields have been found by e.g. Gouveia and Trigo (2008) and Vicente-Serrano et al. (2006), respectively.

In addition to NDVI (Table 1.2), the Vegetation Condition Index (VCI), the Temperature Condition Index (TCI) (Kogan, 1990, 1995) and the Vegetation Health Index (VHI) (Kogan, 1997), have also been successfully considered by several authors for agricultural drought monitoring using remote sensing (Kogan et al., 2005; Bokusheva and Breustedt, 2012; Dalezios et al., 2014). While VCI is based on NDVI and aims to characterize the moisture condition of vegetation, the TCI is based on Brightness Temperature (BT) and characterizes the thermal conditions of vegetation. The VHI, as initial defined by Kogan (1997), is a linear combination of VCI and TCI to consider their combined effect on vegetation health. Nevertheless, some more recent works highlighted the importance of considering optimal weights for the two components of VHI (VCI and TCI), particularly over drylands, which are persistently dominated by VCI in the characterization of the vegetation stress (Bento et al., 2018, 2020).

Besides inducing vegetation stress (Gouveia et al., 2012), drought conditions are also potentially responsible for crop yield losses (Popova et al., 2014; Lesk et al., 2016; Ribeiro et al., 2019a) throughout the world. Among cropping systems, cereals are the more abundant croplands, and rainfed agriculture is the predominant production system worldwide (FAO, 2011). Winter cereals, such as wheat and barley, are two major crops' productions particularly significant in the IP under rainfed conditions (Austin et al., 1998; Vicente-Serrano et al., 2006). As shown by Gouveia et al. (2009), non-irrigated (or rainfed) crops have been critically affected by the extreme drought episodes of 1999 and 2005 in the IP, experiencing a shortening and a delay of the growing season. Nonetheless, the effects of drought on cropping systems vary with several factors, such as month and timescale of the episode (Páscoa et al., 2017b), land-cover type (Gouveia et al., 2016a), vegetation activity (Vicente-Serrano, 2007; Gouveia et al., 2009, 2012), among others. In IP, rainfed cereal yields have shown significant correlations with

NDVI, SPI, PDSI and SPEI (Vicente-Serrano et al., 2006; Gouveia and Trigo, 2008; Páscoa et al., 2017b; Peña-Gallardo et al., 2019), and the dependence of crop yield on climatic droughts is evidenced by the observed decrease of yield in regions with negative correlations (Páscoa et al., 2017b).

In the last decades, the IP has been recurrently affected by droughts (Trigo et al., 2013; Vicente-Serrano et al., 2014). Moreover, future impacts of climate change are also expected to be reflected on drought-related agricultural impacts (Saadi et al., 2015; Leng and Hall, 2019; Yang et al., 2019). The tendency towards warmer conditions during the 21st century is enhancing the occurrence of dry events in the IP (Giorgi and Lionello, 2008; Vicente-Serrano et al., 2014), and subsequently, associated crop yield losses are also expected to increase under climate change scenarios by 2050, particularly the case of rainfed wheat in the Mediterranean (Saadi et al., 2015). In southern Portugal mean wheat yield losses are also prospected to reduce up to 27% during 2051–2080, relative to 1981–2010 (Yang et al., 2019). Although the assessment of climate change agricultural impacts related to future drought conditions were beyond the scope of this thesis, it entails a critical motivation to enhance the knowledge on droughts affecting the IP area with focus on the agricultural impacts.

Other main motivation for a better understanding on drought-related agricultural impacts in the IP relates with the forthcoming risks from compound events, which are high-impact events induced by multiple variables that may not be extreme themselves, but their combination leads to an extreme impact (Zscheischler et al., 2018). The perspective of compound events for understanding extreme impacts meets the most recent guidelines for risk assessment of climate-related consequences, such as the Sendai Framework for Disaster Risk Reduction (Aitsi-Selmi et al., 2015). Diverse types of compound events are recently being proposed, namely preconditioned, multivariate, temporally compounding, and spatially compounding events (Zscheischler et al., 2020). These events are potentially major sources of agricultural damages, namely crop-loss. In particular, the record-breaking French wheat-loss in 2016 has been revealed to be related to the unique combination of a hotter late autumn than usual followed by a wetter than normal spring concurrent with low radiation and potential evapotranspiration (Ben-Ari et al., 2018). In this way, understanding the role played by drought alone or part of a compound event in the agricultural systems over the IP inspired the final steps of the present thesis and enabled new lines of research for future work.

1.1.2 Probabilistic risk analysis

The risk associated with climate extremes is dependent not only on the extreme event itself (which constitutes the hazard), but also on the exposure and vulnerability (Figure 1.2, IPCC, 2012). In the simplest form, the definition of risk is commonly expressed as the combination of the probability of the occurrence of the hazard and the respective consequences (IPCC, 2012):

$$Risk = Probability \times Consequence \quad (1.1)$$

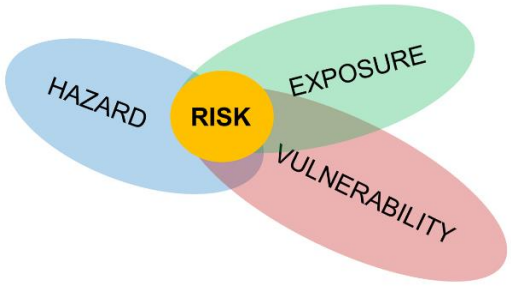


Figure 1.2 - Schematic overview of the components included in risk analysis (hazard, exposure and vulnerability), adapted from IPCC (2012).

According to the IPCC (2012), an increased understanding of the concepts of hazard, vulnerability and exposure is crucial to assess risk. An hazard refers to a categorization of a natural or human induced physical event that may cause a negative effect on people or environment IPCC (2012). The characterization of the drought hazard is performed based on methodologies (e.g. drought indices) which allow for a quantification of a drought event in terms of frequency, duration, intensity and spatial extent (Table 1.1 and Table 1.2). The vulnerability and exposure analysis are more complex tasks. In general terms, vulnerability is the propensity to be adversely affected by the hazard and exposure concerns the presence of people, livelihoods, environmental services and resources, infrastructure, or economic, social, or cultural assets in places that could be adversely affected by the hazard (IPCC, 2012). Both concepts are complementary, and their terminology is dependent on the analysis, objectives and the authors.

In the case of drought-related agricultural impacts, a variety of studies have conceptualized the definitions of vulnerability and exposure in various ways. Dan et al. (2011) assessed the exposure based on planting conditions, vulnerability based on crop drought probability and crop drought tolerance, and drought-resistibility based on irrigation and economic information. On

the other hand, Alonso et al. (2019) characterized vulnerability of the Portuguese agricultural system to drought events considering the three components identified as crucial by the IPCC (exposure, sensitivity and adaptive capacity) based on drought indicators, vegetation indices and soil characterization variables. Alternatively, Popova et al. (2014) has approached agricultural vulnerability to drought using crop models to estimate crop water and irrigation requirements and yield impacts of water stress. According to the IPCC (2012) technological factors, such as irrigation and crop rotation pattern, as well as the development of drought-resistant crops, are important in the case of the agricultural sector.

Prior to the 1990's, most risk studies focused mainly on hazards (IPCC, 2012) and presently, the adopted terminologies are critically dependent on the context of the analysis. Nevertheless, recent research towards the establishment of clear climate risk assessment frameworks stand out that the key aspect of the natural hazards and the associated risks lies in its impacts (IPCC, 2012). In addition, the analytical approaches relative to compound event frameworks generally lie on the analysis of the dependence between the multiple drivers and/or hazards that contribute to the ultimate risk, the extreme nature of the impacts and the probabilistic character of risk analysis (Zscheischler and Seneviratne, 2017; Hao et al., 2018a, 2018c).

Probabilistic risk analysis is a challenging task, and its implementation in its pure form is rarely implemented due to the difficulties associated with the estimation of the probabilities and consequences (IPCC, 2012). In addition, climate change is exacerbating the challenges of implementing probabilistic approaches regarding the risks associated with extreme events (IPCC, 2012). Thus, a joint analysis of the agricultural impacts and the trigger drought conditions considering the multivariate, extreme and regional character of this phenomena is critical to support the decision-making process in the agricultural sector. Additionally, the timescale of the drought episodes is fundamental, as well as the type of crop and the different phenological stages of crop's growth.

In this way, the rationale behind this research focus on an impact-inspired terminology of agricultural drought risk in terms of drought-related crop-losses in the IP. In this thesis the drought-related risk in agriculture is comprehensively defined as the conditional probability of crop-loss under drought conditions, where the hazard is characterized by indicators of drought (Chapters 2, 3 and 4) and heat-stress (Chapter 5), the exposure is characterized by the areas

dominated by rainfed agricultural practices (Chapters 2, 3 and 5) and vulnerability by historical yield data of wheat and barley cropping systems (Chapters 2, 3 4 and 5). In line with the proposed strategies by the IPCC and other international regulatory institutions, the drought-related risk of crop failure is also here investigated as a compound event, which refer to the combination of multiple drivers and/or hazards that contributes to societal or environmental risk (IPCC 2012, Zscheischler et al., 2018, 2020). In particular, the additional agricultural impacts of concurrent drought and heat are compared with the impacts of the individual hazards in Chapter 5.

This thesis focuses on statistical techniques which require lower computation costs than mechanistic modelling in alternative to crop growth models describing crop's biological processes (Hlavinka et al., 2010; Vergni et al., 2015; Capa-Morocho et al., 2016a; Giménez et al., 2016). In this research we make use of four key statistical concepts: correlations, regressions, artificial neural networks and copulas. A common feature to the key statistical concepts employed in this thesis is the assessment of the statistical dependence in data, and how to use those dependencies to perform predictions and probabilistic risk analysis. While the core idea of regression is to fit a model that best fits the data, the linear correlation coefficient is used to characterize the strength of the linear relationship between data. Due to the complexity of the non-linear character of the agricultural systems under drought conditions, artificial neural networks are good alternatives to classical statistical algorithms for prediction (Jiang et al., 2004; Russo et al., 2015a). Chapter 2 of this thesis discusses the use of regression and artificial neural networks techniques for prediction and modelling of the response of crop yield to drought variability.

In the following chapters (Chapter 3 to 5), the concept of conditional probability is employed through the use of copulas to quantify the probability of crop-failure given that drought (Chapters 3 and 4) and/or heat stress (Chapter 5) conditions occur. The application of the copula theory to agrometeorological risk analysis is relatively recent and a promising tool to characterize the multivariate and extreme character of agricultural drought risk (Bokusheva et al., 2016; Madadgar et al., 2017). The core idea of a copula is to link univariate marginal distributions and form the respective joint distribution characterizing the dependence structure between the involved variables. The advantage of using a copula to model dependencies over more classical approaches, like the linear correlation coefficient, is that this last scalar measure

do not disclose important features of the dependence structure, like non-linearities and tail behaviour, whose importance is critical when studying extreme events and the associated risks. Moreover, in comparison to the classic additive models which generally express the effects of explanatory uncorrelated variables on a response variable, copulas consider the dependence patterns in multivariate data, including probable interactions between the explanatory variables. In addition, the proprieties of copulas allow the estimation of conditional probabilities of one extreme event given the occurrence of another extreme event, constituting a valuable and attractive tool in risk analysis as is shown in Chapters 3 to 5 and Appendix A.

1.2 Goals and research objectives

This dissertation is focused on investigating the response of crop yields to drought conditions and developing crop-specific statistical models to estimate the drought-related risks of crop-loss. A combined assessment of the drought hazard and the associated impacts, capturing the multivariate character of drought risk, is the main goal of this research. Ultimately, the additional risk associated with the occurrence of extreme temperatures during droughts completes the research storyline. Throughout this thesis, the following general research questions (RQ) are addressed sequentially:

RQ1. Which are the drought indicators most related with crop yield variability, the relevant timescales of the drought episodes and the moments of the vegetative cycle with highest crop's vulnerability to water stress?

RQ2. What is the skill of several drought indicators (SPEI, VCI, TCI and VHI) in predicting yields before harvest and also events of crop-loss?

RQ3. How to define the risk of drought-related crop-loss in terms of a comprehensively framework?

RQ4. How to characterize the relationship between extreme values of crop yield and drought indicators and how to estimate conditional probabilities of crop-loss?

RQ5. Do conditional probabilities of crop-loss under drought conditions increase when drought events aggravate?

RQ6. Is the risk of drought-related crop failure amplified by the occurrence of hot extremes?

From an operational point of view, this research intends to help contributing to design management options and provide guidance in the decision-making process. In this way, these questions were applied at different regions and/or provinces of the IP and to two different crop species (wheat and barley) to assist stakeholders, such as farmers and insurance companies, in adopting mitigation practices and using supporting tools for agricultural drought risk assessment.

This thesis aims to address the research questions with innovative datasets for drought monitoring and novel and appropriate statistical methods, adding value to the state-of-the-art knowledge on the analysis of drought-related risks in agriculture. From the methodological point of view, the focus on methodologies with ability of capturing joint extremes (here assessed at the two- and three-dimensional spaces) with emphasis on the impacts, unfolds the benefits of using bottom-up approaches and conditional dependencies for risk analysis of extreme events. In this perspective, this thesis presents a novel framework for the assessment of drought-related risks in agriculture and complements conventional analysis in order to contribute to more resilient systems.

1.3 Thesis structure

The thesis is organized around four core chapters sealed by an opening and a closing chapter, adding up to six main chapters. These core chapters are followed by one complementary section, which presents additional work carried out during the PhD and relates marginally with the thesis plan (Figure 1.3).

The structure of the thesis is as follows:

- **Chapter 1** presents the background on the topic and states the research goals

- **Chapter 2** investigates the response of crop yields to drought conditions over major agricultural areas using Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN) (Ribeiro et al., 2019b)

- **Chapter 3** designs a statistical model based on copula theory to address the dependence between crop yields and drought conditions using joint distributions and further estimates of conditional probabilities of crop-loss (Ribeiro et al., 2019c)

- **Chapter 4** defines drought-related risks in agriculture as the conditional probability of occurring crop-losses under drought conditions and maps the risk at the province level of the IP for different drought severity conditions (Ribeiro et al., 2019a)

- **Chapter 5** investigates the trivariate joint distribution of crop yields, maximum temperature and precipitation, and the role played by the interaction between hot and dry extremes in amplifying the agricultural impacts of the individual hazards (Ribeiro et al., 2020 preprint)

- **Chapter 6** summarizes the main conclusions of the conducted research, the main achievements and deliverables of the PhD, and enunciates several future lines of research

- **Appendix A** extends the knowledge on compound dry and hot conditions to the occurrence of drought-related hot summers in the IP (Ribeiro et al., in review)

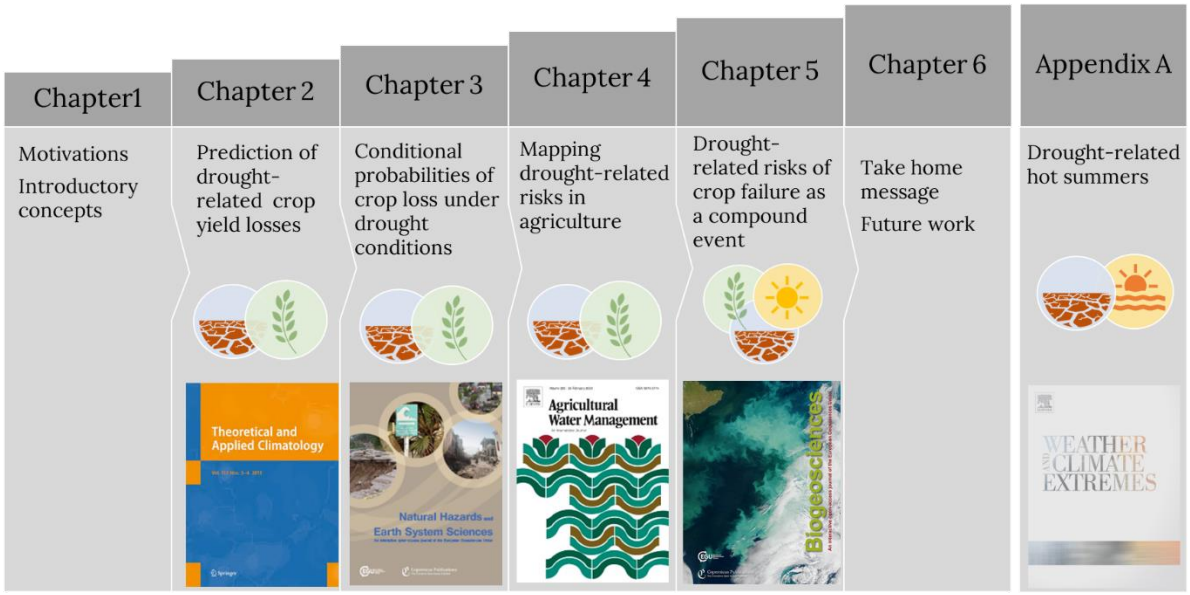


Figure 1.3 – Schematic overview of the thesis structure.

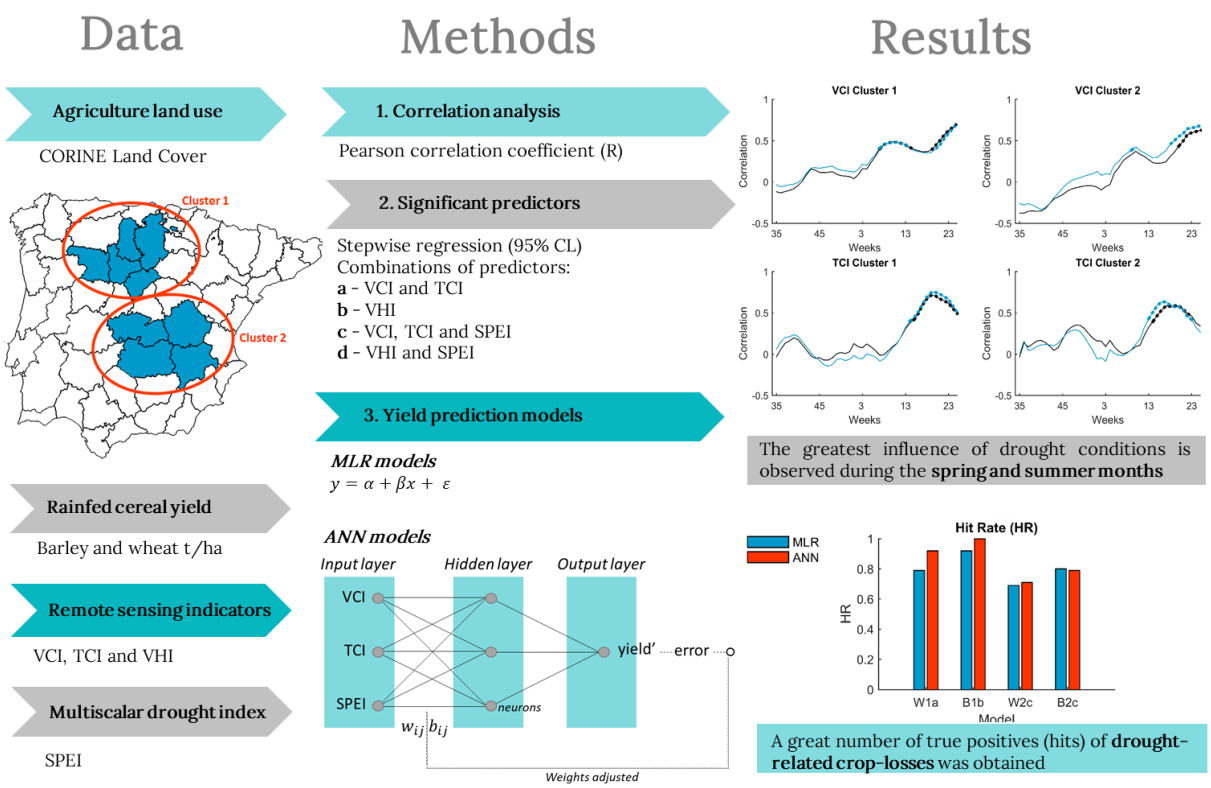
Chapter 2

Modelling drought-related yield losses in Iberia using remote sensing and multiscalar indices

***Ribeiro, A.F.S.**, Russo, A., Gouveia, C.M., Páscoa, P. (2019). Modelling drought-related yield losses in Iberia using remote sensing and multiscalar indices, Theoretical and Applied Climatology, 136, 203-220, https://doi.org/10.1007/s00704-018-2478-5*

***Andreia F. S. Ribeiro contributions:** preparation of the data (crop yields, remote sensing and SPEI); statistical analysis and interpretation of the results (correlation analysis, stepwise regression, multiple linear regression models, artificial neural networks, contingency tables); production of all the figures and coordination of the manuscript writing.*

Graphical abstract



Highlights

- Two clusters of provinces dominated by rainfed agricultural practices were identified in the Iberian Peninsula (IP)
- Drought events over the IP, namely during 1992, 1995 and 2005, are associated with low values of yield at both clusters and cereals (wheat and barley)
- The correlation between drought indicators and crop yields is generally stronger during the crop stages in which the vegetation is photosynthetically more active (spring and summer), rather than the earlier moments of plants life cycle (autumn/winter)
- Correlation analysis suggests that crop yields reduce with vegetation moisture depletion (low values of VCI) during early-spring and with high temperatures (low values of TCI) during early-summer
- Regression models' outputs indicate stronger relationships between the remote sensing indices and crop yields in the northern cluster, in contrast to stronger relationships between SPEI and crop yields in the southern cluster
- The overall performance of the artificial neural networks and regression models shows a good ability to estimate the occurrence of crop-loss
- The statistical performance of the artificial neural networks and regression models is similar

Abstract

The response of two rainfed winter cereal yields (wheat and barley) to drought conditions in the Iberian Peninsula (IP) was investigated for a long period (1986-2012). Drought hazard was evaluated based on the multiscalar Standardized Precipitation Evapotranspiration Index (SPEI) and three remote sensing indices, namely the Vegetation Condition (VCI), the Temperature Condition (TCI) and the Vegetation Health (VHI) Indices. A correlation analysis between the yield and the drought indicators was conducted, and Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) models were established to estimate yield at the regional level. The correlation values suggested that yield reduces with moisture depletion (low values of VCI) during early-spring and with too high temperatures (low values of TCI) close to the harvest time. Generally, all drought indicators displayed greatest influence during the plant stages in which the crop is photosynthetically more active (spring and summer), rather than the earlier moments of plants life cycle (autumn/winter). Our results suggested that SPEI is more relevant in the southern sector of the IP, while remote sensing indices are rather good in estimating cereal yield in the northern sector of the IP. The strength of the statistical relationships found by MLR and ANN methods is quite similar, with some improvements found by the ANN. A great number of true positives (hits) of occurrence of yield-losses exhibiting Hit Rate (HR) values higher than 69% was obtained.

Keywords: Drought; Rainfed cereals yield; Remote sensing indices; Standardized Precipitation Evapotranspiration Index (SPEI); Iberian Peninsula

2. Modelling drought-related yield losses in Iberia using remote sensing and multiscalar indices

2.1 Introduction

Crop production is directly affected by the weather and climatic conditions (Capa-Morocho et al., 2016b). In the Mediterranean basin, natural vegetation in general and crop production in particular has always been affected by large natural climate variability (Grasso and Feola, 2012; Gouveia et al., 2017b; Páscoa et al., 2017b), and is expected to continue to be affected in the future (Nguyen et al., 2016). Particularly, seasonal changes in precipitation and temperature and their seasonal variability affects crop production, especially in regions where crops are highly dependent on precipitation (Ruiz-Ramos and Mínguez, 2010). As a consequence of long-term influence of precipitation and temperature on crop production, drought is a major cause of unexpected crop failure (Wilhelmi and Wilhite, 2002; Wu and Wilhite, 2004; Vicente-Serrano et al., 2006; Li et al., 2009; Falco et al., 2014; Lesk et al., 2016). In a climate change context, one of the key aims in the agricultural sector for the next few decades will be the mitigation of the risk associated with drought-related crop-losses (Li et al., 2009; Ferrise et al., 2011; Capa-Morocho et al., 2016a).

A significant part of the Iberian Peninsula (IP) countries' economies and landscape is linked to agriculture. In 2014, the IP had more than 26 million hectares of harvested area, and about 2% of each IP countries' gross domestic product (GDP) came from the agriculture sector (FAO, 2015). Among agricultural crops, winter cereals such as wheat and barley are two major world crop productions (FAO, 2014) particularly significant in the Mediterranean regions, and the growing of these cereals under rainfed conditions is dominant in the IP countries (Austin et al., 1998; Vicente-Serrano et al., 2006).

Presently, the increase of the frequency of occurrence of drought events in the IP (Vicente-Serrano et al., 2014; Páscoa et al., 2017b) and the close relationship between cereal yield and drought conditions in the Iberian territory is pointed out by several authors (Vicente-Serrano et al., 2006; Iglesias and Quiroga, 2007; Páscoa et al., 2017b). Austin et al. (1998) have shown a strong dependence of wheat and barley on seasonal rainfall in Spain, and the response of winter

cereals in IP to the widely used precipitation-based Standard Precipitation Index (SPI) have also been demonstrated in several works (Vicente-Serrano et al., 2006; Iglesias and Quiroga, 2007; Hernández-Barrera and Rodríguez-Puebla, 2017). Moreover, and aside from rainfall variability, drought severity in southwestern Europe is being reinforced by enhanced evaporative demand due to an increased temperature scenario (Trigo et al., 2013; Vicente-Serrano et al., 2014). Hernández-Barrera and Rodríguez-Puebla (2017) have found wheat yields to be declining in Spain due to warming climate conditions, and according to Ferrero et al. (2014), maize yield in Spain using rainfed systems may be at risk as heat waves will increase in intensity, frequency and duration. Consequently, under the scope of climate change, a sustainable agricultural management of rainfed crops requires reliable estimations of the drought impacts using diverse drought indicators at various spatial and temporal scales.

To include the effect of evapotranspiration on drought monitoring, the Standardized Precipitation Evaporation Index (SPEI) was proposed (Vicente-Serrano et al., 2010) and is now widely used (Vicente-Serrano et al., 2014; Gouveia et al., 2017b; Zampieri et al., 2017). In the IP, rainfed cereal yield have shown significant correlations with SPEI varying with several factors, such as month and time scale of the dry episode (Páscoa et al., 2017b). Atmospheric patterns, such as the North Atlantic Oscillation (NAO) have also shown significant relationships with wheat yield in the IP (Gouveia and Trigo, 2008; Capa-Morocho et al., 2016b).

In addition to the hydro-meteorological influence, the recent advances of remote sensing have strongly contributed to the agricultural sector (Rojas et al., 2011; Kogan et al., 2015a; Van Hoolst et al., 2016). The widely used Normalized Difference Vegetation Index (NDVI) was reported to be strongly (negative) correlated to the winter wheat yield over the southern part of Portugal (Alentejo) (Gouveia and Trigo, 2008) and north of Spain (Ebro valley) (Vicente-Serrano et al., 2006). Moreover, remote sensing indices based on NDVI and Brightness Temperature (BT) have also been successfully considered by several authors for modelling agricultural productivity (Dalezios et al., 2014; Kogan et al., 2015a, 2015b; Bokusheva et al., 2016), including the Vegetation Health Index (VHI) (Kogan, 1995), the Vegetation Condition Index (VCI) (Kogan, 1990) and the Temperature Condition Index (TCI) (Kogan, 1995).

An important step towards developing strategies to mitigate agricultural drought risk is the establishment of models for estimating crop yield under drought influence (Vicente-Serrano et

al., 2006; Kogan et al., 2015a; Mishra et al., 2015). In mechanistic modelling, crop yield is estimated by equations describing the relationships between complex biophysical variables and crop growth, requiring a high degree of input data (Paredes et al., 2014, 2016; Giménez et al., 2016). On the other hand, empirical modelling makes use of statistical relationships between yield data and predictor variables, representing rather well larger scale impacts of drought conditions (Vicente-Serrano et al., 2006; Kogan et al., 2015a; Matsumura et al., 2015). Despite the lack of detailed representation of crop's biophysical interactions, empirical modelling is computationally easier and have lower computation costs than mechanistic modelling, and the results are considered good (Ferrise et al., 2011; Estes et al., 2013). Results found by Ferrise et al. (2011) suggested a high level of correspondence between a mechanistic model of durum wheat in the Mediterranean with empirical model's results. The authors successfully used Artificial Neural Network (ANN) models to reproduce the results of a wheat yield mechanistic model output by using mean spring temperature and precipitation (Ferrise et al., 2011).

The applications of ANN have been increasing in the recent past for modelling and prediction on environmental studies (Morid et al., 2007; Russo et al., 2013; Le et al., 2017) and have proved to add significant improvements to traditional statistical modelling, such as Multiple Linear Regression (MLR) models, namely in the case of crop yield modelling (Jiang et al., 2004; Matsumura et al., 2015).

The purpose of the current work is to model, through the application of MLR and ANN techniques, the influence of drought conditions in rainfed winter cereal yields (wheat and barley) over the major agricultural areas in the IP, examining the potential of combining remote sensing indices (VCI, TCI and VHI) with a multiscalar drought indicator (SPEI). The results presented in this paper constitute a first step towards the development of an agricultural drought risk model for the IP, and may contribute to assist final users and insurance companies with some guidance on decision making process.

2.2 Data and Methods

2.2.1 Rainfed cereal yields and land cover in Iberia

Agricultural drought especially affects the growing of crops under rainfed conditions (Páscoa et al., 2017b) making data on agricultural land use and harvested yields key factors in agricultural drought risk reduction. Hence, maps of land cover information and data on two major rainfed crops in the IP (wheat and barley) were analysed over the Iberian territory. In IP the precipitation regime is marked by a strong variability (Martin-Vide and Lopez-Bustins, 2006; Muñoz-Díaz and Rodrigo, 2006; Martins et al., 2012), hence there is a high probability of occurrence of droughts and the agricultural activities are particularly prone to its effects. The highly variable precipitation regime in space and time over the IP is strongly associated with the geographic diversity of the peninsula, like the orography, and the influence of diverse circulation weather patterns (Cortesi et al., 2014). The spatial patterns of rainfall in the IP exhibit strong gradients, with higher values in the northwestern sector and lower values in the southeastern sector, and most of precipitation is concentrated between October and May (Belo-Pereira et al., 2011). In addition to the lack of rain, drier conditions in the summer are enhanced by high temperatures during the summer in the IP (Vicente-Serrano et al., 2014). The spatial heterogeneity of vegetation dynamics in the IP is pronounced, with predominance of the vegetation classes with the maximum of vegetation greenness in spring (Gouveia et al., 2017b). According to the classification by Gouveia et al. (2017), the spatial distribution of vegetation clusters exhibits a northwestern-southeastern gradient: Temperate Oceanic – Mediterranean Oceanic – Mediterranean dry. The vegetation behaviour in the IP ecosystems is mainly driven by the precipitation regimes (Gouveia et al., 2008), being expressed in the vegetative cycle of the winter crops: sowing usually occurs between October and November and the harvest occurs during June and July of the following year (Gouveia and Trigo, 2008; Capa-Morocho et al., 2016b).

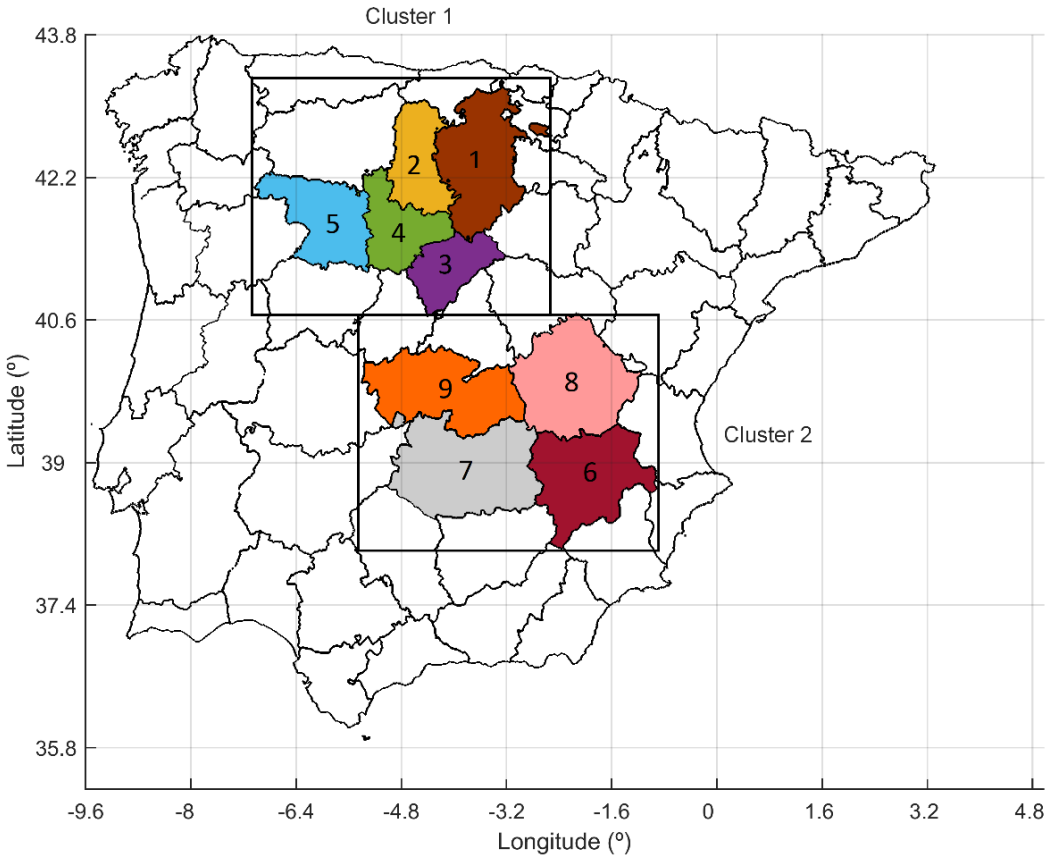


Figure 2.1 - Selected clusters of provinces correspondent to the agricultural drought prone areas. Cluster 1 provinces: Burgos (1), Palencia (2), Segovia (3), Valladolid (4) and Zamora (5). Cluster 2 provinces: Albacete (6), Ciudad Real (7), Cuenca (8) and Toledo (9).

Annual production (tons, t) and total area (ha) of barley and wheat crops were obtained from the Portuguese National Statistics Institute (INE) and the Spanish Agriculture, Food and Environment Ministry, for the regions of Portugal and the provinces of Spain, respectively. Annual crop yield time-series were calculated as the ratio between the collected crop's production and harvested area during the period of 1986-2012 (Páscoa et al., 2017b). The year 1986 corresponds to the year when the crop yield time-series in Portugal started to be aggregated at the regional (and not only at district level as until 1985) level, as they are available in Spain, and therefore considered as the beginning of the analysis (Páscoa et al., 2017b). The crop yield anomalies were computed by removing the crop yield time series linear trend, in order to exclude non-climatic factors (Gouveia and Trigo, 2008; Páscoa et al., 2017b).

The pixels corresponding to rainfed cereal crop areas were identified considering the non-irrigated arable land classification from the more recent CORINE Land Cover map (CLC 2012)

which is a standard procedure (Vicente-Serrano et al., 2006; Gouveia et al., 2011, 2016a; Atzberger et al., 2014; Blauhut et al., 2016). As not all provinces are strongly dominated by agricultural practices, a selection of the major rainfed agricultural areas in the IP is required (Figure 2.1). The provincial clusters selected for the present analysis have been determined according to three criteria: 1) the provincial land use is dominated by agricultural practices, i.e. more than half of the pixels at each province correspond agricultural areas; 2) the agricultural areas are dominated by rainfed crops, i.e. more than half of the agricultural areas correspond to non-irrigated arable land; 3) the provinces are contiguous and non-isolated. Selecting provincial clusters provides the advantage of estimating a short number of models for a larger number of provinces. In this way, we intend to estimate the best model for each cereal over each cluster, applicable to more than one province.

2.2.2 Remote sensing and multiscalar indices

With the aim of evaluating the response of the rainfed winter cereal yields (wheat and barley) to the regional drought conditions, drought hazard was evaluated based on the multiscalar drought index SPEI and the remote sensing indices VCI, TCI and VHI. The potential of modelling cereal crop in the IP based on these drought indicators, considering different combinations of the possible predictors (as will be described later), is one of the goals of the present study.

The above mentioned remote sensing indices are based on NDVI and BT, given that green vegetation reflect visible and emit thermal solar radiation. The VCI and TCI are mathematically expressed by weekly NDVI and BT values, respectively, relative to their minimum and maximum limits and further normalized relative to their amplitude interval (Equation 2.1 and Equation 2.2). Mathematical expressions of VCI and TCI were first introduced by Kogan (1990 and 1995), respectively, where a detailed description of the indices calculation was provided. The VCI and TCI characterize the moisture and thermal conditions of vegetation, respectively, and the VHI (Equation 2.3) is assumed as an average of the two in order to consider their combined effect of vegetation health (Kogan, 1997).

$$\mathrm{VCI} = 100 \times \frac{\mathrm{NDVI} - \mathrm{NDVI_{min}}}{\mathrm{NDVI_{max}} - \mathrm{NDVI_{min}}} \quad (2.1)$$

$$\mathrm{TCI} = 100 \times \frac{\mathrm{BT_{max}} - \mathrm{BT}}{\mathrm{BT_{max}} - \mathrm{BT_{min}}} \quad (2.2)$$

$$\mathrm{VHI} = \left(\frac{\mathrm{VCI} + \mathrm{TCI}}{2}\right) \quad (2.3)$$

The values of VCI, TCI and VHI vary from 0 to 100, and index values below 40 are indicative of drought conditions (Kogan, 2001). The reason for employing these remote sensing indices in the present study, instead of the popular NDVI, is the inclusion of the thermal component (BT) and their ability to consider ecosystem changes in terms of fluctuations between the maximum and minimum values of NDVI and BT. Accordingly with their definition (Kogan, 1997), low values of VCI indicate vegetation stress due to lack of water content and low TCI values correspond to vegetation stress due to high temperatures.

The weekly global maps of VCI, TCI, and VHI were retrieved at 4km spatial resolution from NOAA's ftp server (ftp://ftp.star.nesdis.noaa.gov/pub/corp/scsb/wguo/data/VHP_4km/geo_TIFF), during 1985-2012. The reason of the inclusion of weekly data for 1985 is because the plant life cycle of the cereals harvested in 1986 starts in the autumn/winter of the year before, in this case 1985. Missing week values were substituted by the climatological value of each week, and the analysis was performed between the week 35 (approximately the beginning of September of the year n – 1) and 25 (approximately the end of June of the year n), comprising the major crop life cycle moments: pre-sowing and sowing (autumn/winter), vegetative phase (winter/early spring), reproductive phase (middle of spring), stage of formation and maturation of the grain (end of spring), and beginning of crop harvest (early summer). The spatial averages of VCI, TCI and VHI were computed for each provincial cluster and used for further cereal yield modelling.

One of the aims of the present study is to discuss the utility of the remote sensing indices for cereal yield modelling, assessing the relative contribution of the moisture and thermal term and the further combination with the additional information of the drought index SPEI. Thus, the monthly drought index SPEI gridded values, with spatial resolution of 0.5°, were computed

based on precipitation and temperature values from the Climate Research Unit (CRU TS3.21). The SPEI computation uses the monthly difference between precipitation (P) and potential evapotranspiration (PET) as shown in Equation 2.4, where D provides a simple measure of the water deficit for the analysed month at different time-scales.

$$D = P - PET \tag{2.4}$$

A log-logistic distribution was used, as suggested by Vicente-Serrano et al. (2010), and the Hargreaves method was considered for the estimation of the reference evapotranspiration (Beguería et al., 2014). A discussion of several computing options for the use of SPEI is provided by Beguería et al. (2014). The spatial averages of SPEI at the time-scales 1-12 months were computed for each provincial cluster from January to June. The use of a variety of time-scales (1-12 months) incorporates the memory of the respective past months, which does not happen with remote sensing indices, and for this reason the SPEI data considered for the analysis covers approximately the period between the crop growth vegetative phase to the harvest (January to June). In other words, the SPEI period in analysis don't include the typical months of pre-sowing and sowing because their drought conditions are intrinsically considered in the medium and longer time-scales of the SPEI intervals (4 to 12 months).

2.2.3 Linear correlation analysis

Having identified the cluster of provinces more exposed to agricultural drought, a correlation analysis is conducted to assess the linear relationships between the winter cereal yields and the drought indicators (remote sensing and multiscalar indices) in terms of the Pearson correlation coefficient (R) (Wilks, 2006). Statistical significant evidence is assessed with a 95% significance level.

The moments of the vegetative cycle of the highest crop's requirements to moisture and thermal conditions are assessed in terms of VCI and TCI, respectively. The relationships between the cereal yield and the VHI indicate the impacts of the combined effect of water and heat stress during the crop growth cycle. In addition, the winter cereal yield response to each time scale of drought occurrence is assessed based on the multiscalar drought index SPEI during the development stages of the cereals.

2.2.4 Selection of significant predictors and their possible combinations

The range of predictors encompasses three remote sensing indices (43-week intervals for each) and one multiscalar drought index (6 months (January to June) by 12 time-scales = 72 SPEI intervals) for each of the provincial clusters. The time scales and months of SPEI, together with the weeks of VHI, VCI and TCI better related with wheat and barley yield were chosen based on stepwise regression (95% confidence level). The stepwise regression algorithm carries out an exhaustive search and generates a subset of predictors which together have the largest contribution to the variability of each cereal yield in each provincial cluster (predictands). For each provincial cluster and each winter cereal (wheat and barley), stepwise regression models are performed based on the moisture and thermal components (VCI and TCI) separately from models based on the VHI, to avoid collinearity since VHI is a combination of both VCI and TCI. Subsequently, stepwise regression models combining SPEI with the remote sensing indices (VCI+TCI+SPEI and VHI+SPEI) are performed to evaluate the relative contribution of the remote sensing indices and the further combination with the multiscalar index for the simulation of the variability of winter cereal yield.

2.2.5 Cereal yield estimation models

After the selection of the significant predictors, the standardization of both dependent and independent variables is performed by computing the z-scores for further statistical modelling (Wilks, 2006). Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) techniques are applied for modelling the wheat and barley yields at the provincial clusters. The reason for the application of a non-linear methodology in addition to the classical MLR models, is to discuss the use of alternative promising tools, such as the ANN (Morid et al., 2007; Russo et al., 2015a), to simulate the complexity of the non-linear character of the agricultural systems under drought conditions (Jiang et al., 2004; Matsumura et al., 2015).

In MLR, the functional relationship between the predictand (cereal yield) and the predictors (previously statistically selected) can be described by means of the intercept and the slope of the regression line, usually called regression coefficients. The regression coefficients are estimated by minimizing the sum of the squared differences between the observations of cereal yield and the regression line (Wilks, 2006).

ANN are mathematical models inspired by the behaviour of the human nervous system, composed by several layers and respective neurons. In this study, a simple three-layer structure was adopted with one input layer, one hidden layer and one output layer. The input variables corresponding to the statistically significant predictors are forced by the weight and bias, which alter the initial information at the neurons, and then pass the combined information to the next layer and consequently reach the output value of the simulated cereal yield (target). The ANN training that updates the weights and bias on each cycle was here performed according to the Levenberg-Marquardt backpropagation method and considering the same statistical significant predictors (input variables) as the MLR models. Different architectures were examined in which different number of neurons in the hidden layers between 1 and 4 were considered. In order to compare the different architectures a fixed seed was considered for the initial random weights. The use of a second hidden layer was tested but it was found to be redundant. The number of neurons in the input layer corresponds to the number of selected predictors for each target (wheat and barley yield at cluster 1 and 2). For the single node in the output layer a linear transfer function was considered, and for each hidden neuron a log-sigmoid function was considered to account for the non-linear behaviour.

The MLR and ANN model's performance is assessed in terms of leave-one-out cross-validation, obtaining unbiased estimations by avoiding overfitting associated to the models, which occurs when the same data is used for the fit and for the performance assessment. The leave-one-out cross-validation assesses how well the model performs by successively using a small set of observations from the original sample for validation, and the remaining observations as the training data. In other words, in the present work, one observation is successively removed from the total sample for the model's fit (training data), and the left-out observation is used for validation (validation data). This procedure ensures that every data is used for training and validation independently, since the model's performance is assessed on independent data not considered on the fit. This approach is commonly used and appropriate for cases which have a low number of samples (Wilks 2006), as is the present study. To support the robustness of the leave-one-out cross-validation scheme, the results of explained variance in terms of adjusted coefficient of determination (R^2_{adj}) are analysed with and without cross-validation mode ($R^2_{adj\_no\_cv}$). The adjusted R^2_{adj} is an unbiased R^2 considering the finite sample and the number of predictors used as input for the MLR and ANN models. Other widely used

accuracy measures are also considered to evaluate the performance of the linear and non-linear methods, such as the Root Mean Squared Error (RMSE) and the Skill Score based on the RMSE (SS_{RMSE}, Equation 2.5). The total deviance of simulated values from observed values is assessed in terms of the RMSE, and the SS_{RMSE} (Equation 2.5) is used in this paper considering persistence (the previous year yield value) as a reference model.

$$SS_{RMSE} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{y} - y_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_{i+1} - y_i)^2} \times 100 \qquad (2.5)$$

Having statistically modelled the standardized anomalies of wheat and barley yields at the regional level by MLR and ANN techniques, the potential of the modelled cereal data for prediction of crop-losses is assessed. Here, crop yield loss is defined as values of standardized yield anomaly below zero, indicating the years when harvested cereal crops is below the mean value. The MLR and ANN model's performance regarding the loss of crop yield (yield anomaly < 0) is assessed in terms of contingency tables and the associated categorical scores (Wilks, 2006): Frequency Bias (FB), Success Ratio (SR), Hit Rate (HR), and False Alarm Rate (FAR). The score FB describes the ratio of the estimated and observed events and measures the ability of the models to underestimate (FB<1) or overestimate (FB>1) the occurrences of crop-loss. For example, models with FB > 1 indicate that occurrences of crop-losses were modelled more often than they occur and FB = 1 indicates that the model is unbiased. The score SR describes the ratio between the hits and the estimated events and gives information about the likelihood of a crop-loss, given that it was predicted by the model. The HR and FAR scores correspond respectively to the rate of correct forecast of crop-loss (proportion of occurrences which are hits) and the rate of wrong forecast of crop-loss (proportion of non-occurrences which are false alarms).

2.3 Results

2.3.1 Cereals and drought indicators during low yield years

Two clusters of provinces dominated by rainfed agricultural practices are identified (Figure 2.1), according to the criteria described in section 2.2.1. Both clusters are in Spain, approximately in the regions of Castilla-Léon and Castilla-La Mancha. The northern provincial cluster (cluster 1) includes 5 provinces (Zamora, Valladolid, Palencia, Segovia and Burgos) and the southern provincial cluster (cluster 2) includes 4 provinces (Toledo, Cuenca, Ciudad Real and Albacete). Figure 2.2 shows the spatial averages of wheat and barley yield from 1986-2012, computed for each provincial cluster. The corresponding trends and detrended time-series (yield anomalies) are also illustrated. The temporal evolution of the yield anomalies shows low values (i.e. below the 25th percentile) during drought episodes over the IP, particularly during the events which took place during 1992, 1995 and 2005 (years associated with low yield at both clusters and cereals). Figure 2.2 shows that the more recent events of 1995 and 2005 experienced yield anomalies more negative in the southern sector of the IP (cluster 2 - Figure 2.2 bottom panel), while the year of 1992 exhibited yield anomalies more negative in the northern sector (cluster 1 - Figure 2.2). Overall, the temporal evolution of wheat and barley yield anomalies is similar at both provincial clusters, although the respective productions and total crop areas are quite distinct at the province level (not shown).

Figure 2.3 and Figure 2.4 show drought severity during the individual low yield years of 1992, 1995 and 2005, based on the spatially distributed averaged values over each cluster of the remote sensing indices (VCI, TCI and VHI, Figure 2.3) and the SPEI (Figure 2.4). Figure 2.3 shows the weekly values of VCI, TCI and VHI from the week 35 (year n – 1) to week 25 (year n), corresponding approximately to the period between the sowing and harvesting of the winter cereals, i.e. between September of previous year to June of harvest year. According to the remote sensing indices (Figure 2.3), there is little or no evidence of drought conditions (values below 40) during the first growth stages (until January) for both clusters during the 3 considered years, particularly featuring cold autumn/winter weeks based on the high values of TCI (low values of BT) during 1992. Similar values of TCI are found in 2005 during intermediate growth stages of the cereal life cycle (week 11), indicating a spring with low temperatures. On the other

hand, in 1995 during the intermediate and final growth stages (more evident in cluster 1), the VCI values are indicative of favourable moisture conditions (NDVI increase), in contrast with hotter conditions found by low values of TCI (BT increase). Nevertheless, there is almost no evidence of drought based on VHI, during 1995 and 2005 in cluster 1 (except June). The highest number of drought weeks recorded by the VHI (also coincident with low values of VCI and TCI) are found in cluster 1 during 1992, and in cluster 2 during 1995 and 2005. While the onset of drought conditions in 1992 (cluster 1) is experienced during vegetative growth stages (winter), 1995 and 2005 (cluster 2) show less favourable conditions slightly later. This feature is in accordance with the regions of Iberia that were more affected by drought in 2005 which was more intense in southern Iberia (Gouveia et al., 2012).

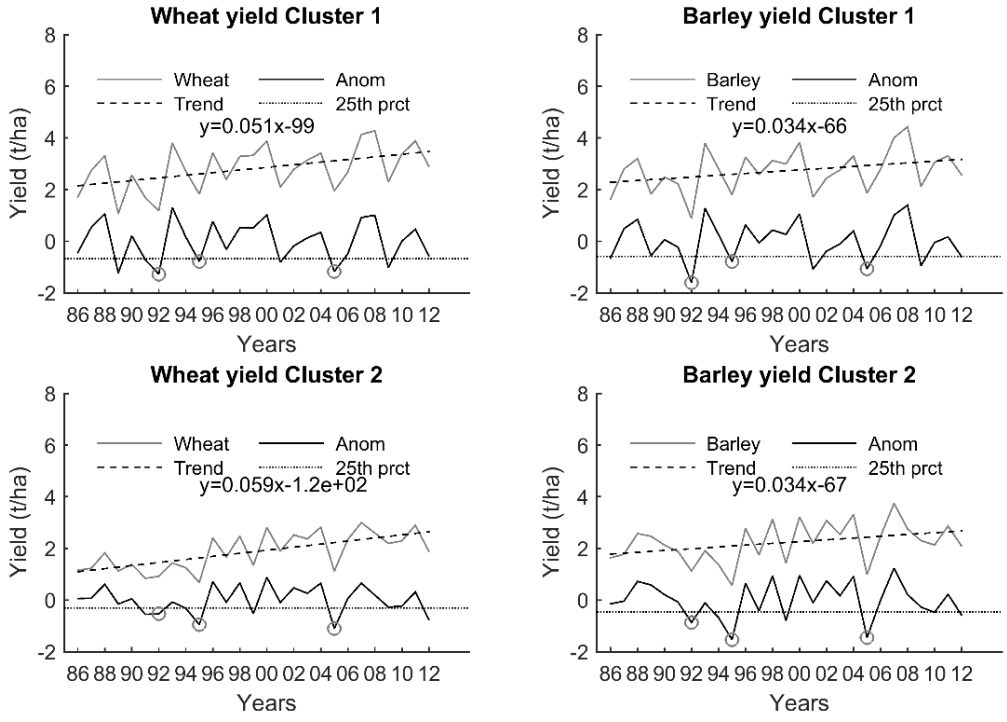


Figure 2.2 - Wheat and barley yields (grey lines), trends (dashed lines and respective equations), anomalies (black lines) and 25th percentile of yield anomalies (dotted line) during the period 1986-2012 over the two selected provincial clusters. The common years associated with low yield anomalies (i.e. below the 25th percentile) are denoted by a circle: 1992, 1995 and 2005.

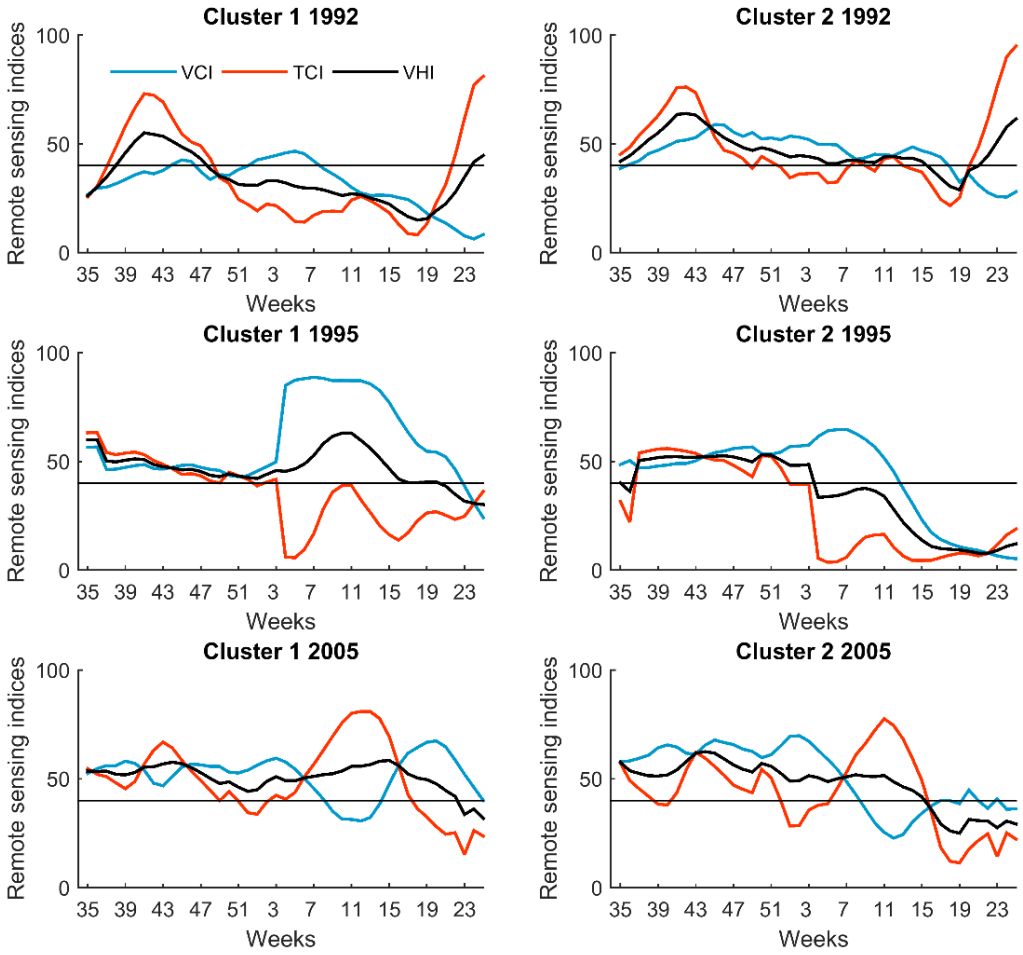


Figure 2.3 - Weekly values of spatial averages of VCI (Vegetation Condition Index), TCI (Temperature Condition Index) and VHI (Vegetation Health Index) between the week 35 (beginning of September of the year n-1) and 25 (end of June of the year), during the low yield years of 1992 (top panel), 1995 (middle panel) and 2005 (middle panel) at the cluster 1 (left) and cluster 2 (right). Values below 40 indicate drought conditions.

Figure 2.4 shows the monthly values of the SPEI at the different time-scales (1-12 months) between January and June, corresponding approximately to the period between the vegetative growth stage and harvesting. In 1992 the overall pattern shows values of SPEI indicating drought or near normal conditions, namely for the first months of the year and for longer times scales. On the other hand, spring months do not present a marked pattern, showing a tendency to wet conditions, in particular in June for cluster 2. This feature may be associated with the non-droughts conditions based on TCI and VHI (low temperatures and favourable vegetation conditions), in contrast with drought conditions displayed by the VCI (moisture stress) (see Figure 2.3 top panels). These finding suggests that despite the presence of favourable conditions according to SPEI, TCI and VHI, the greenness of vegetation was shallow (low values of VCI) during the final growth stages. Moreover, cluster 1 does not show a clear pattern of drought conditions in 1995 and 2005. In fact, during these years the drought, as obtained by SPEI, is

evident only in April (June) during 1995 (2005). On the other hand, extreme drought conditions, accordingly with SPEI, were observed in 1995 and 2005 over cluster 2, being stronger in May (June) during 1995 (2005). These results are also in accordance with the ones obtained using the vegetation indices (Figure 2.3).

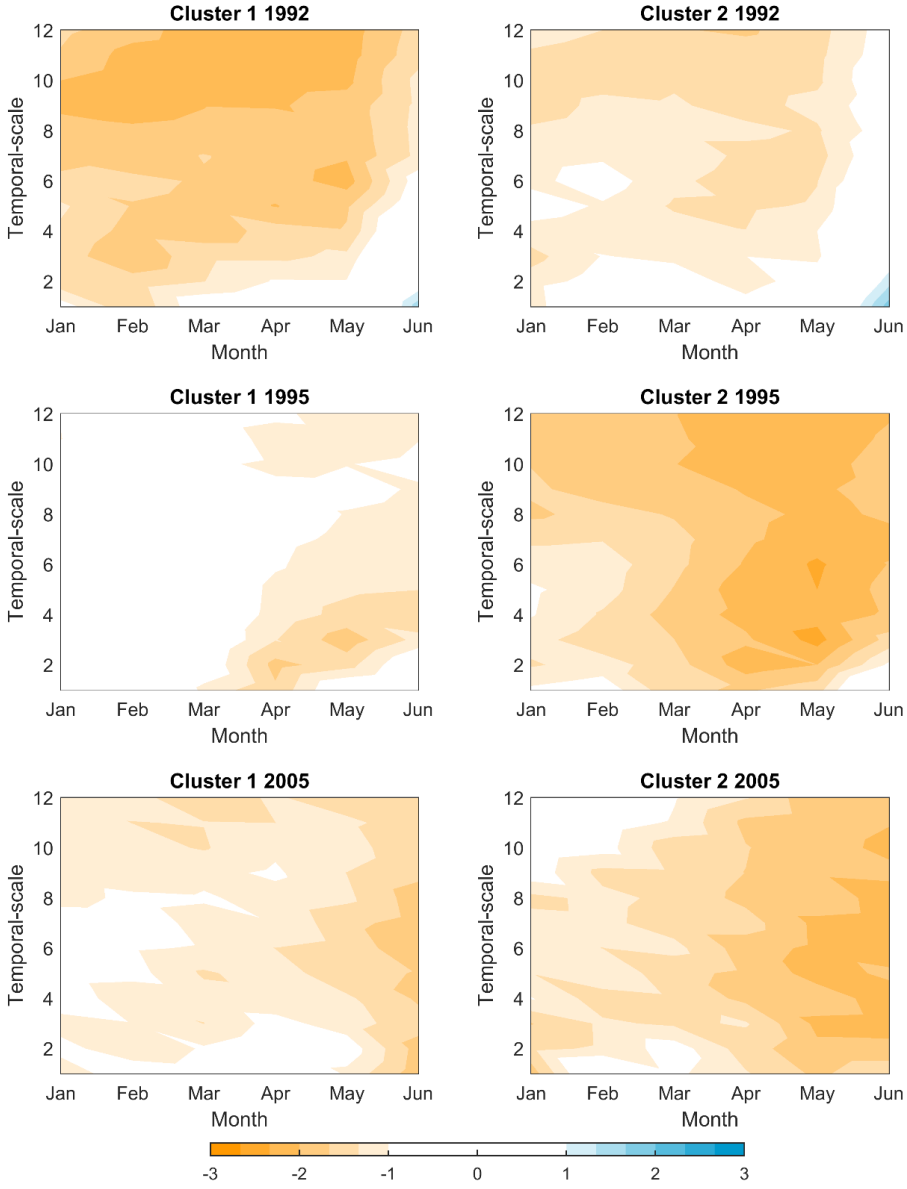


Figure 2.4 - Monthly values of spatial averages of SPEI at time scales from 1 to 12 months (y axis) between January and June during 1992 (top panel), 1995 (middle panel) and 2005 (bottom panel) at the cluster 1 (left) and cluster 2 (right). Values between 1 and -1 correspond to near normal conditions, and values below -1 and above 1 indicate dryness and wetness, respectively.

In general, a good agreement is found between the higher values of negative yield anomalies (Figure 2.2) and the drought-affected weeks according to the remote sensing indices and the months of SPEI (Figure 2.3 and Figure 2.4). In 1992, detrended yield time-series display higher

negative anomalies in cluster 1 rather than in cluster 2 (Figure 2.1), in accordance with drier conditions suggested by the remote sensing indices and SPEI in cluster 1 as well (Figure 2.3 and Figure 2.4). Similarly, 1995 and 2005 display more pronounced negative anomalies of yield (Figure 2.2) and drier conditions (Figure 2.3, Figure 2.4) in cluster 2. In conclusion, negative yield anomalies followed by dry conditions in 1992 were more pronounced in the northern sector of the IP (cluster 1), while the same conditions in 1995 and 2005 were more pronounced in the southern sector (cluster 2).

2.3.2 Relationships between cereal yield and drought indicators

To investigate the strength of the relationship between the winter cereals crop yield and the remote sensing indices, and to identify the moments of the vegetative cycle of the highest crop's requirements to moisture (VCI) and thermal conditions (TCI), a correlation analysis was performed (Figure 2.5). Figure 2.5 shows the correlation coefficients between the winter cereals yield (barley and wheat) and the remote sensing indices over the two agricultural provincial clusters from week 35 to week 25. Generally, VCI, TCI and VHI display low correlations during the first growth stages of both rainfed cereal (during autumn and begin of winter) and a sharp increase from the intermediate growth stages to the harvest time over both provincial clusters. This feature is consistent with Figure 2.3, which shows no evidences of drought conditions until further growth stages during the low yield years (1992, 1995 and 2005) according to the VCI, TCI and VHI. In the same way, correlation values suggest that the greatest influence of the remote sensing indices is observed during the spring and summer months, corresponding to the moments in which the vegetation is photosynthetically more active.

Moreover, between the late winter and the early summer, VHI and VCI correlation values are statistically significant, whereas TCI significant correlations are found between early spring and early summer. This aspect points out that while water stress (VCI) on vegetation exhibits stronger correlations during early-spring (late February and early March approximately), heat stress (TCI) shows stronger correlations slightly afterwards during the latter growth stages (from the 14th week (April) onwards). In other words, Figure 2.5 suggests that crop yield decline is associated with moisture depletion on vegetation (low VCI) during early-spring and with high temperatures (low TCI) close to the harvest time. The correlations obtained with the

VHI are generally stronger than the VCI and TCI, and exhibit a peak during late-spring with potential impacts on the maturation of both wheat and barley grains. The temporal evolution of the correlation values is very similar in both cereals throughout the crop life cycle, with some barley correlation values slightly stronger in cluster 2.

The crop response to each SPEI time-scale (1 to 12 months) was evaluated for each cereal at each cluster, approximately from the vegetative phase to the harvesting moment (January to June). The results are illustrated in Figure 2.6 for the 1-, 3-, 6-, 9- and 12-month time-scales, representative of the shorter (1 and 3), medium (6) and longer time-scales (9 and 12). Similarly to Figure 2.5, SPEI displays lower correlations during the vegetative growth stages of both rainfed cereals, over the two provincial clusters (Figure 2.6). As a matter of fact, the SPEI exerts major influence during the springer months (April to June), particularly at the shorter time-scales (1 and 3-months), corresponding to intermediate and final growth stages. May exhibits the strongest correlations in all cases. At the time-scales of 6-, 9- and 12-months the correlation values during the winter are more pronounced than for the shorter time-scales (1 and 3-months), and the difference between the different seasons is not so evident. In cluster 2, the correlation values of SPEI with 6- and 9-month time-scales are statistically significant during the whole growth cycle, while in the other cases most of the statistical significance is found during intermediate growth stages (spring). At all time-scales, the number of statistically significant correlations is higher in cluster 2 (southern sector) rather than in cluster 1 (northern sector). Moreover, the impact of SPEI on cereal yield in cluster 2 is registered earlier than on cluster 1, considering all the temporal scales.

In general, the correlations in cluster 1 during spring in Figure 2.5 reach stronger values than in Figure 2.6, suggesting stronger relationships between remote sensing indices and cereal yield in the northern sector, rather than SPEI. On the other hand, cluster 2 exhibits more months with statistically significant correlations in Figure 2.6 (particularly at 6- and 9-month time-scales) than in Figure 2.5, suggesting stronger relationships between SPEI and cereal yield in the southern sector (cluster 2).

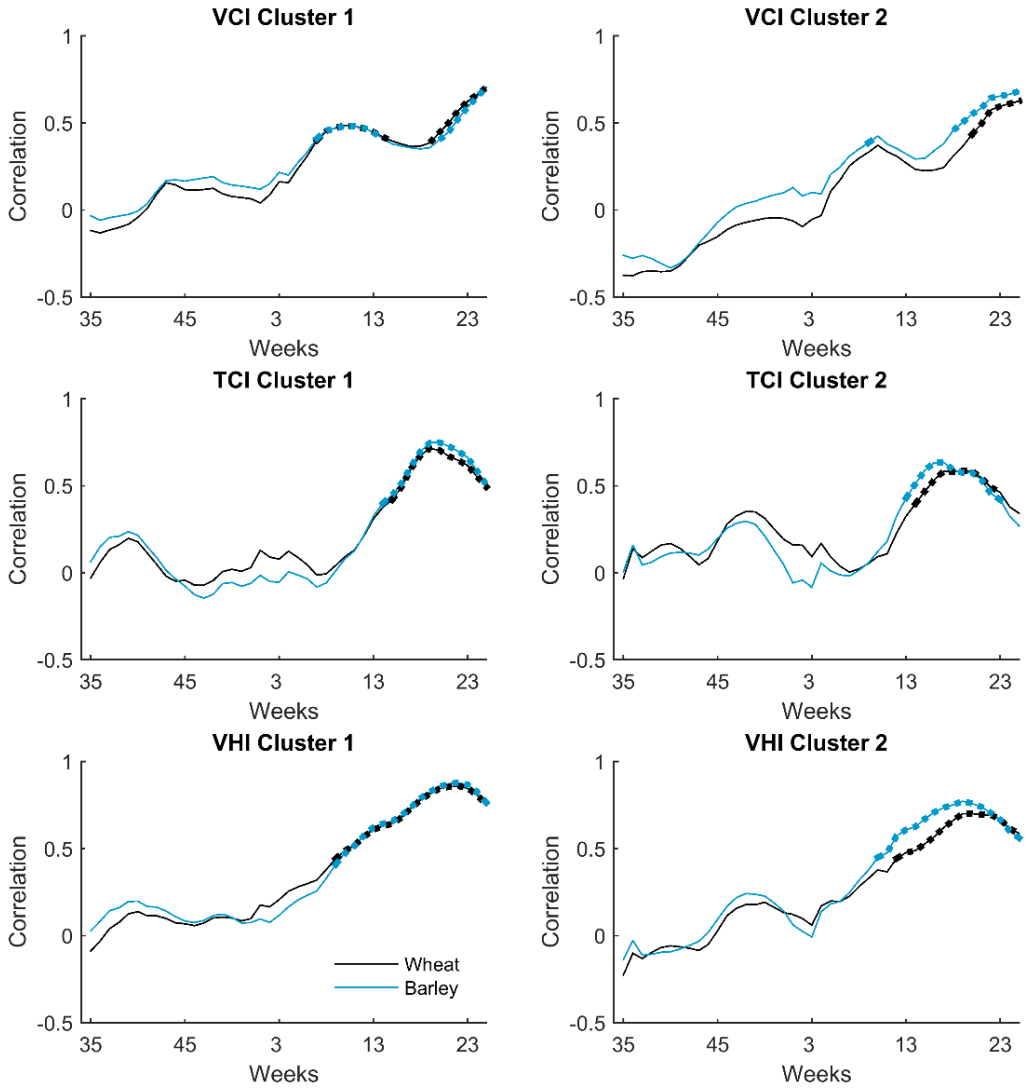


Figure 2.5 - Correlations between the weekly values of VCI (Vegetation Condition Index), TCI (Temperature Condition Index), VHI (Vegetation Health Index), and the wheat yield (full line) and the barley yield (dashed line) in cluster 1 (left) and cluster 2 (right), between 1986-2012. The significant correlations at 95% level of confidence are marked with a dot over the line.

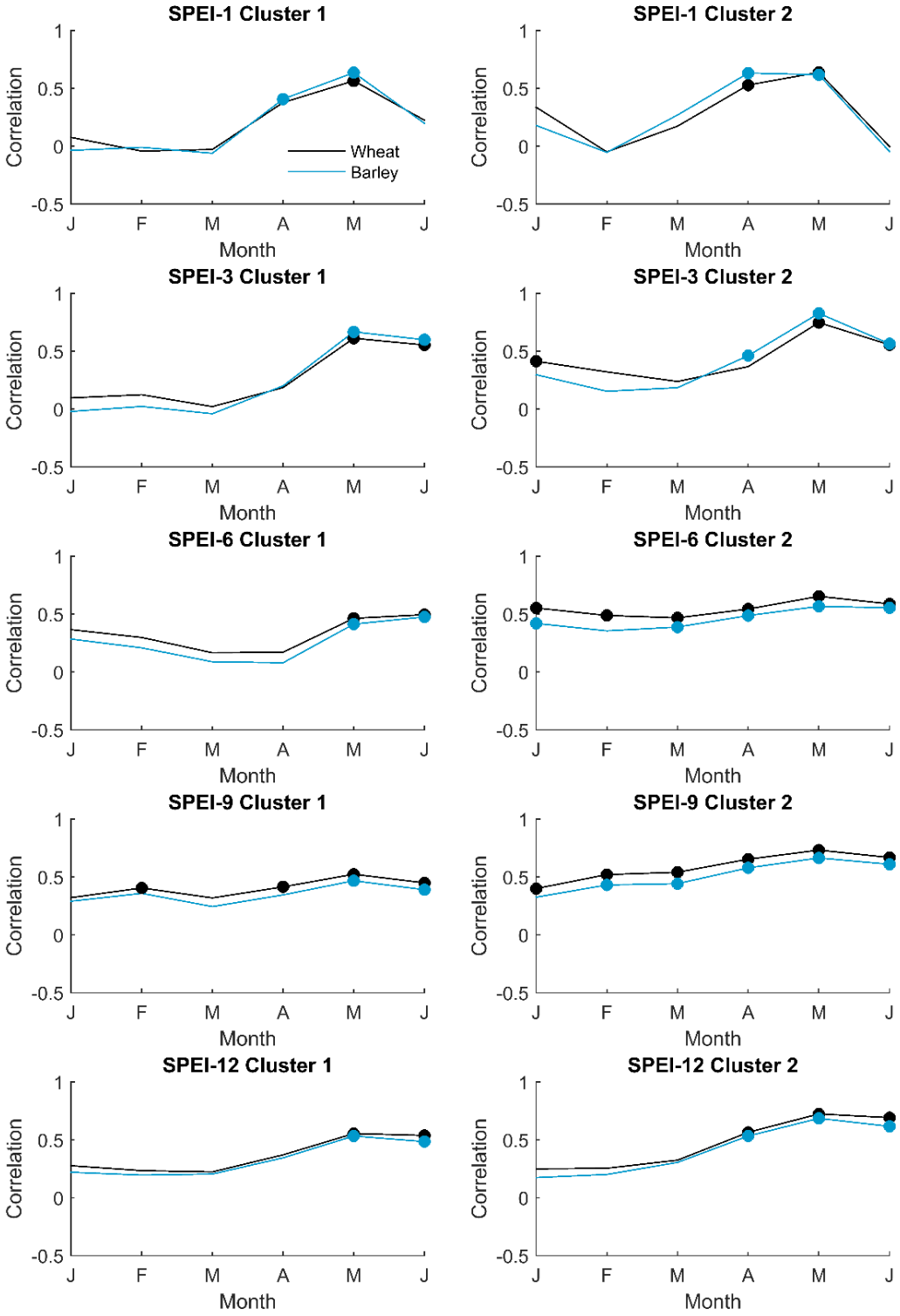


Figure 2.6 - Correlations between average SPEI and wheat yield (full line) and barley yield (dashed line) in cluster 1 (left) and cluster 2 (right), between January and June of 1986-2012. The results are illustrated for 1, 3, 6, 9 and 12-month time-scales, representative of the shorter, medium and longer time-scales. The significant correlations at 95% level of confidence are marked with a circle.

2.3.3 Statistical significant predictors/inputs

The correlation analysis between yield and the drought indicators pointed out significant temporal differences of the drought impact, and pointed to different moments of the vegetative cycle when the crops are more vulnerable to drought conditions (Figure 2.5 and Figure 2.6). Therefore, the redundant information should be removed to find the time scales and months of SPEI, together with the weeks of VHI, VCI and TCI most suitable to accurately estimate the cereal yield. The statistical significant predictors were chosen based on stepwise regression (95% confidence level). Table 2.1 shows the selected predictors for each of the 4 combinations of predictors, resulting on 11 different subsets of input variables for the MLR and ANN models.

Each resulting model nomenclature (Table 2.1) refers to the target cereal species (letter "W" for wheat and letter "B" for barley), the respective provincial cluster (clusters 1 and 2), and the possible combination of predictors (VCI and TCI - "a"; VHI - "b"; VCI, TCI and SPEI – "c", VHI and SPEI – "d"). For example, the model W1a refers to the wheat yield at cluster 1, based on the statistically significant weeks of VCI and TCI, and model W2c refers to the wheat yield at cluster 2, based on the statistically significant time scales and months of SPEI in addition to the better related weeks of VCI and TCI.

In accordance with the correlation analysis in Figure 2.6, the results from stepwise regression (Table 2.1) indicate that the inclusion of the drought index SPEI in the pool of possible predictors is only significant in the cluster 2 for both cereals. In the case of the cluster 1, the predictor selection chooses the same variables for the pair of models "a" and "c", and the same for "b" and "d". In other words, the inclusion of SPEI information is redundant in cluster 1. In consequence, only models based on VCI and TCI together (W1a and B1a based on late spring weeks 18, 20, 21, 23) and VHI (W1b and B1b based on mid-winter and late-spring weeks 50, 1 and 22) are performed in cluster 1. The SPEI of February, April, May and June display significant influence at cluster 2, when SPEI is included in the predictors' pools. In fact, in models W2c and B2d the remote sensing indices weeks are removed by the stepwise regression, remaining only SPEI information to estimate the cereal yields.

The selected remote sensing indices weeks suggest a predictive power based on the autumn/early-winter period and mid-spring/early-summer weeks (Table 2.1). Between the week 18 (~mid-April) and 25 (~mid-June) 10 predictors (remote sensing indices) are selected,

and between the week 35 (~early-September) and 1 (early-January) 9 predictors are selected. Between January and mid-spring only the SPEI of February with 5-months' time-scale is selected as predictor. In comparison with cluster 1, the predictor selection in cluster 2 selects a larger number of winter and late autumn variables, particularly in the case of barley cereal (Table 2.1). The models B2a, B2b and B2c select the earlier week values of the three predictors (vegetation indices in late autumn, winter and spring), and model B2d selects SPEI of February, April and June, similarly to models W2c and B2c. Only the barley model B2c selects VCI and SPEI together as statistical significant predictors.

Finally, it is important to stress that most of the models select 2 or 3 predictors, whereas the model B2c is the one with highest number of predictors (n=5). On the other hand, only one model (W2b) chooses only 1 predictor (VHI).

Table 2.1 - Results of the stepwise regression at the 95% confidence level for the selection of the statistical significant predictors of wheat (white columns models) and barley (gray columns models) yields (p indicates the number of predictors select for each model). The numbers of the remote sensing indices correspond to the respective weeks selected to predict the cereal yield, and the numbers of SPEI correspond to the respective month and time-scale selected by each model.

Model name / **Drought indicator**	**W1a**	**W1b**	**B1a**	**B1b**	**W2a**	**W2b**	**W2c**	**B2a**	**B2b**	**B2c**	**B2d**
VCI	21	-	18	-	35 25	-	-	40 52 25	-	40 51	-
TCI	23	-	20	-	-	-	-	-	-	-	-
VHI	-	1 22	-	50 22	-	20	-	-	43 49 19	-	-
SPEI	-	-	-	-	-	-	Feb-5 Apr-1 May-12 Jun-9	-	-	Feb-5 Apr-1 Jun-6	Feb-5 Apr-1 Jun-5
p	2	2	2	2	2	1	4	3	3	5	3

2.3.4 MLR and ANN models

The overall performance of the MLR leave-one-out cross-validation is shown in Table 2.2 in terms of the statistical measures described in section 2.2.5, for the 11 possible models. The results indicate that model B2c presents the highest values of performance explaining 85% of the variance of barley yield in cluster 2, based on the VCI, TCI and SPEI. The values of the explained variance without cross-validation ($R^2_{adj\_no\_cv}$) are slightly less conservative than the

values obtained by cross-validation (R^2_{adj}) in all models, supporting the robustness and reliability of the models. While models W2a and B2a explain less 11% of the variance with cross-validation, the remaining models explain less than 10% with cross-validation reaching a low of 5% less by the models W1b and B1b.

Concerning the wheat cereal in cluster 1 (Table 2.2), the R^2_{adj} and RMSE of the models W1a and W1b display the same values (73%), while the W1a values of SSRMSE display marginally higher percentage of performance against persistence. The barley cereal in cluster 1 denotes higher explained variance (76%) and lower RMSE considering the VHI as predictor (B1b). The rainfed cereals in cluster 1 exhibit the strongest linear relationship considering the late-spring weeks of VCI and TCI as predictors in the case of wheat (W1a), and the mid-winter and late-spring weeks of VHI in the case of barley (B1b).

Regarding the cluster 2, the use of SPEI in the predictors' pool shows an added value in combination with VCI and TCI (W2c and B2c). Models W2c and B2c display the highest values of explained variance (71% and 85% respectively) and 69% and 78% of the skill against persistence. While barley at cluster 2 displays the strongest linear relationship based on a remote sensing index (VCI) and SPEI together, the model W2c consists only of SPEI values (VCI and TCI are not significant predictors). In comparison with models without the multiscalar drought index in cluster 2, the inclusion of SPEI reduces the importance of the VCI and VHI values (TCI is not a significant predictor in any model in cluster 2). In the case of the model B2c, the VCI during week-25 (proximate to the harvest) used in model B2a is "replaced" by the SPEI predictors.

Table 2.2 - Summary of the regression equations and the overall performance of the MLR leave-one-out cross-validation. The models of wheat (white columns models) and barley (gray columns models) at cluster 1 and 2. For each cereal and cluster, the models with highest performance are denoted by a *.

Model name	***p***	**Summary of Regression Equations**	$R^2_{adj\_no\_cv}$	R^2_{adj}	***RMSE***	SS_{RMSE} **(%)**
W1a*	2	$W1 = 0.67*VCI_{21} + 0.76*TCI_{23}$	0.79	0.73	0.49	69.31
W1b	2	$W1 = 0.24*VHI_{1} + 0.88*VHI_{22}$	0.78	0.73	0.49	68.87
B1a	2	$B1 = 0.47*VCI_{18} + 0.82*TCI_{20}$	0.76	0.69	0.52	66.86
B1b*	2	$B1 = 0.23*VHI_{50} + 0.91*VHI_{22}$	0.81	0.76	0.47	70.41
W2a	3	$W2 = -0.39*VCI_{35} + 0.63*VCI_{25}$	0.50	0.39	0.74	53.42
W2b	1	$W2=0.70*VHI_{20}$	0.47	0.39	0.75	52.40

W2c*	4	$W2 = 0.94*SPEI_{2-5} + 1.05*SPEI_{4-1} – 0.53*SPEI_{5-12} – 0.32*SPEI_{6-9}$	0.80	0.71	0.49	69.23
B2a	3	$B2 = – 0.63*VCI_{40} + 0.52*VCI_{52} +0.65*VCI_{25}$	0.67	0.56	0.61	60.99
B2b	3	$B2 = – 0.34*VHI_{43}+0.49*VHI_{49}+0.76*VHI_{19}$	0.68	0.58	0.60	61.69
B2c*	5	$B2 = –0.39*VCI_{40} + 0.34*VCI_{51}+1.07*SPEI_{2-5} +0.91*SPEI_{4-1}–0.84*SPEI_{6-6}$	0.91	0.85	0.34	78.20
B2d	3	$B2 = 1.14*SPEI_{2-5}+0.86*SPEI_{4-1}–0.78*SPEI_{6-6}$	0.83	0.75	0.46	70.74

Table 2.3 shows the performance of the ANN models in terms of the same statistical measures as those used for MLR models. For sake of simplicity, the presented ANN results are shown based on the most suitable ANN architectures according to the skill against persistence prediction (SS_{RMSE}). In general, a good performance is observed considering between 1 and 5 hidden neurons. The model B2c also presents the highest values of performance but explaining slightly less variance (84%) than the B2c MLR model (85%). Similar to Table 2.2, the ANN statistics supports the robustness of the models using cross-validation (Table 2.3). However, models W2a and W2b significantly decrease the explained variance by using cross-validation, while in the remaining models the difference with and without cross-validation is similar to the observed by MLR models.

The statistics present in Table 2.3 indicate that 5 ANN models (W1a, B1a, B1b, W2c and B2b denoted by a ‘) improve the MLR results (Table 2.2). Similar to the linear regression statistics, the models W1a, B1b, W2c and B2c display the strongest relationships explaining 85%, 83%, 73% and 84% of the variance in the case of ANN models, against 73%, 76%, 71% and 85% in the case of MLR models respectively (denoted by a * in Table 2.2 and Table 2.3). Hereafter results are presented only for the models W1a, B1b, W2c and B2c since they present the best performance for each cereal in each cluster considering both MLR and ANN techniques (Table 2.2 and Table 2.3). Except for the case of B2c, the highest performance models are slightly improved using ANN techniques. The overall good performance of the models is illustrated in Figure 2.7, which shows the time-series of the cereal observations in each cluster, together with the respective estimations using ANN and MLR methods. The considerable similarities between the two techniques in model B2c is well shown in the bottom panel of Figure 2.7, while some differences are observed in the other models.

Table 2.3 - Summary of the ANN architectures according to the maximum SS_{RMSE} value and the overall performance of the respective ANN leave-one-out cross-validation. The architectures indicate the number of neurons in input, hidden and output layer respectively. For each cereal and cluster, the models with highest performance are denoted by a *. The models which performance of the ANN techniques improves the MLR results (Table 2.2) are denoted by a '.

Model	*p*	Architecture	$R^2_{adj\_no\_cv}$	R^2_{adj}	*RMSE*	SS_{RMSE} (%)
W1a*'	2	2-3-1	0.93	0.85	0.36	77.30
W1b	2	2-1-1	0.80	0.71	0.51	67.85
B1a'	2	2-4-1	0.90	0.80	0.42	73.30
B1b*'	2	2-4-1	0.89	0.83	0.39	75.20
W2a	2	2-3-1	0.65	0.36	0.76	52.04
W2b	1	1-2-1	0.57	0.12	0.90	43.20
W2c*'	4	4-2-1	0.89	0.73	0.47	70.34
B2a	3	3-1-1	0.67	0.49	0.66	57.76
B2b'	3	3-3-1	0.89	0.75	0.46	70.56
B2c*	5	5-1-1	0.91	0.84	0.36	76.96
B2d	3	3-1-1	0.83	0.74	0.47	69.92

A summary of contingency table results of the occurrence of crop yield losses (standardized yield anomaly < 0) is presented in Figure 2.8, comparing the performance of the MLR and ANN techniques of the models W1a, B1b, W2c and B2c. The results show that the models W1a, B1b and B2c based on ANN, and B1b based on MLR slightly overestimate the yield losses, while the remaining models are almost unbiased (FB~1). Generally, all models predict a great number of true positives (hits) of occurrence of crop-losses exhibiting HR values higher than 69%. Except for B2c, the ANN models display values of HR higher than MLR models, estimating more occurrences of crop-loss. The SR values indicate that in the case of B1b and W2c, the likelihood of crop-loss occurrence, given that it was estimated by the model, is higher based on ANN rather than on MLR techniques. In comparison with wheat, the barley models display slightly higher values of SR and HR. The cereal models in cluster 2 display the lower values of SR, HR and higher values of FAR, in comparison with the cluster 1.

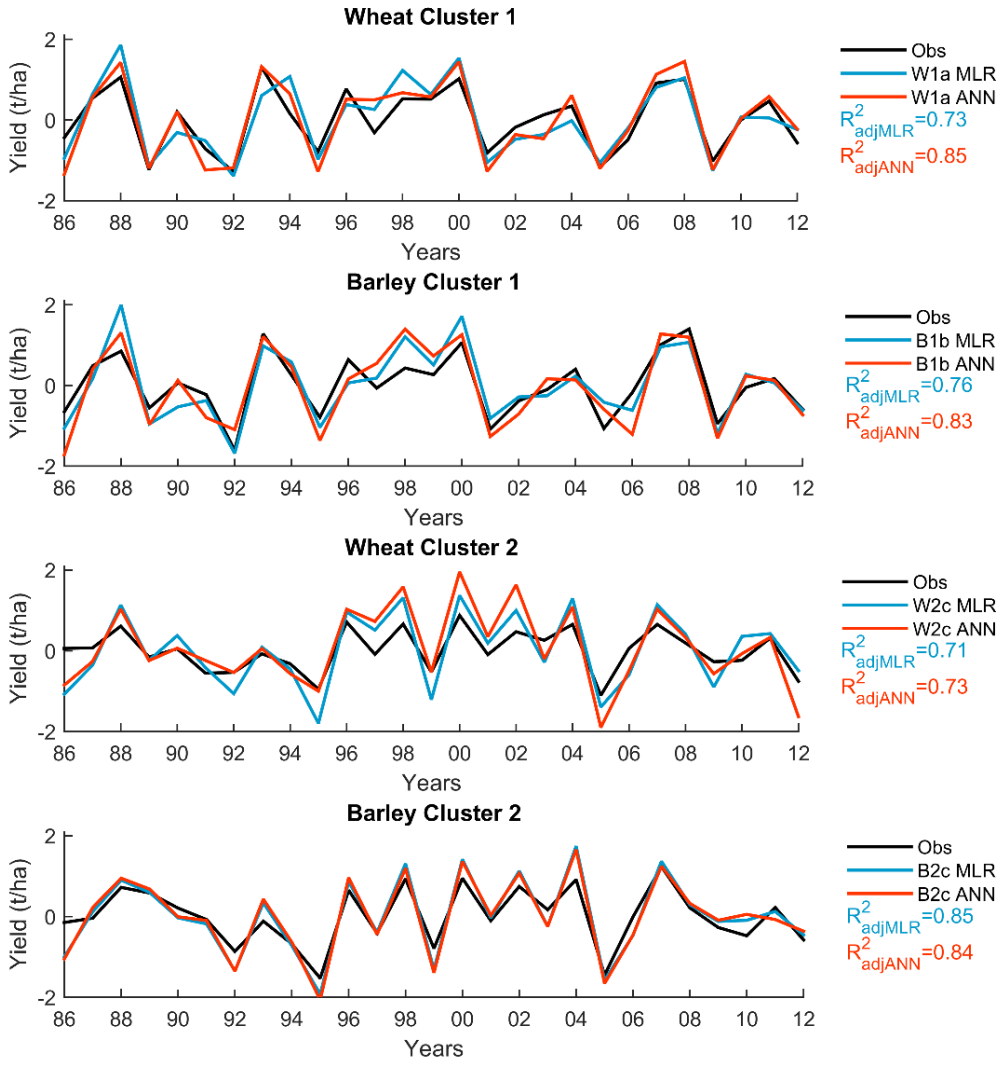


Figure 2.7 - Wheat and barley time-series of observations (full line) from 1986-2012 in clusters 1 (top two panels) and 2 (bottom two panels) and respective statistical estimations using MLR (dotted line) and ANN (dashed line) methods with the strongest statistical relationships (W1a, B1b, W2c and B2c).

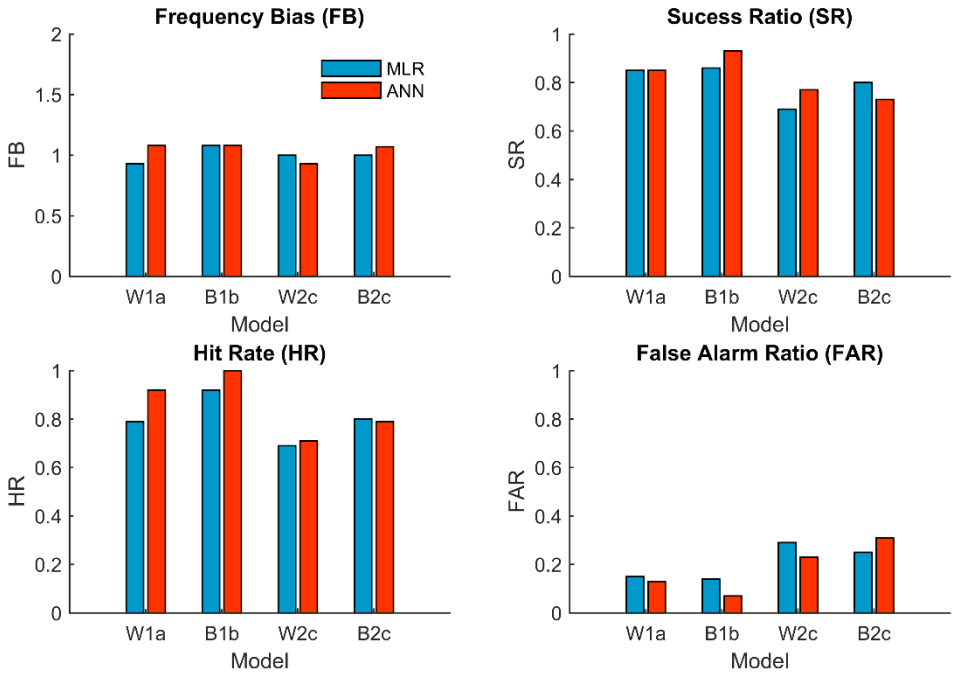


Figure 2.8 - Summary of contingency table results of the occurrence of crop yield losses (standardized yield anomaly < 0) in terms of Frequency Bias (FB), Success Ratio (SR), Hit Rate (HR), and False Alarm Rate (FAR), based on MLR (white bars) and ANN (black bars) methods of the models W1a, B1b, W2c and B2c.

2.4 Discussion and conclusions

This work aimed to assess the influence of drought conditions in agricultural yields over the IP, considering remote sensing (VCI, TCI and VHI) and multiscalar (SPEI) drought indices as predictors of rainfed cereal yields. The exposure analysis performed in this work allowed for the identification of distinct geographical areas in the IP exposed to agricultural drought, according to the use of dryland for agriculture. In a different way from the criteria applied in the present work, Hernández-Barrera and Rodríguez-Puebla (2017) have also specified two different regions in the IP by applying a cluster analysis based on wheat yield data variability. Other approach followed by Iglesias and Quiroga (2007) selected 5 sites representing the major rainfed and irrigated agricultural regions of Spain. In this work, the analysis of exposure to agricultural drought in terms of dryland allowed for the study of more than one cereal growing in rainfed conditions, and proved to be rather suitable for wheat and barley. Moreover, the aggregation of provinces with similar percentage of arable land allowed the estimation of a few number of models suitable for a larger number of provinces.

We found that the spatial averages of wheat and barley computed for each cluster exhibited low values of yield anomalies during the years of 1992, 1995 and 2005 (Figure 2.2), coinciding with main drought events that affected the IP (García-Herrera et al., 2007; Andrade and Belo-Pereira, 2015). As a matter of fact, the drought conditions identified with the remote sensing and multiscalar indices (Figure 2.3 and Figure 2.4) are coincident with the low yield anomalies: drier conditions were found in the northern cluster in 1992 (anomalies more negative in cluster 1 according to Figure 2.2), while drier conditions (anomalies more negative in cluster 2 according to Figure 2.2) were found in the southern cluster in 1995 and 2005. The temporal evolution of the drought hazard during the individual low yield years of 1992, 1995 and 2005 (Figure 2.3 and Figure 2.4) and the correlation analysis (Figure 2.5 and Figure 2.6) also suggested minor influence of drought conditions during the initial growth stages (autumn/winter), and greatest influence during the intermediate and final growth stages (spring/summer), corresponding to the moments in which the vegetation is photosynthetically more active. Stronger relationships with NDVI were also found by Vicente-Serrano et al. 2006 during flowering stages of wheat and barley yield in north-east Spain (Middle Ebro valley).

Given the importance of assessing crop's vulnerability to dry conditions at different stages of plant development, we also looked for the highest crop's requirements to moisture (VCI) and thermal (TCI) conditions at different moments of the vegetative cycle (Figure 2.5). The correlation values between crop yield and remote sensing indices suggested that crop yield reduces with moisture depletion (low values of VCI) during early-spring (and enhance with water content increase) and with too high temperatures (low values of TCI) close to the harvest time (and improve with temperature decrease). This highlights the importance of both water content and air temperature for cereals productivity and the advantage of combining the contributions of moisture and thermal conditions using the remote sensing indices. The effects of water stress and high temperatures during middle growth stages of the crop life cycle is in accordance with previous studies (García del Moral et al., 2003; Iglesias and Quiroga, 2007; Ferrise et al., 2011).

The use of remote sensing and multiscalar indices (Figure 2.3 and Figure 2.4) allowed analysing the vegetation responses to drought conditions over large regions and at different time-scales of drought occurrence. The dominant time-scales at which drought influences the crop yield correspond to longer time-scales (6 to 12-months) throughout January to June, and a

pronounced impact is verified during the springer months (April to June) at the shorter time-scales (1 to 6-months) (Figure 2.5 and Figure 2.6). These results are in accordance with previous work performed by the authors, which stress the stronger impact of longer timescales and identify spring as the dominant season of winter cereal yield dependence on drought conditions based on SPEI, particularly in Spain (Páscoa et al., 2017b).

Spatial differences were also pointed out by the correlation analysis and the statistical modelling, suggesting that in comparison with cluster 1 (northern sector), cluster 2 (southern sector) is impacted by dry conditions beforehand (Figure 2.5 and Figure 2.6), in accordance with the geographical location and respective climate variability of the provinces. According to Rodriguez-Puebla et al. (1998), the spatial patterns of the precipitation regime in Spain exhibit strong gradients, with higher values in the northwestern sector and lower values in the southeastern sector. In addition, the southern sector of Iberia have been exceptionally affected by severe drought events, particularly during the recent episode of 2004/2005 (García-Herrera et al., 2007; Gouveia et al., 2009, 2012), which was also a year with higher negative yield anomalies in cluster 2 (Figure 2.2, Figure 2.3 and Figure 2.4).

Significant regional differences were also found considering the potential of combining the multiscalar and remote sensing approaches. Correlation analysis results from Figure 2.5 and Figure 2.6 suggested stronger relationships between remote sensing indices and cereal yield in the northern sector (cluster 1), and stronger relationships between SPEI and cereal yield in the southern sector (cluster 2). In agreement, the results of the stepwise regression for significant predictors selection (Table 2.1) suggested that the inclusion of the drought index SPEI in the possible predictors pools is only significant in the cluster 2 for both cereals. These findings propose that the southern sector crop yield is better inferred from the drought index SPEI information, rather than remote sensing indices, and sensitive to balance between precipitation and evapotranspiration. On the contrary, cluster 1 models suggest strong dependence of the health of the vegetation, and only predictors based on VCI, TCI and VHI are selected (Table 2.1). Kogan et al. 2015b and Kogan et al. 2004 have also performed accurate predictions of crop yield based on VCI, TCI and VHI for Russia and China, respectively, and have suggested the potential of using the remote sensing of vegetation health to assess weather-related crop-losses.

The combined use of remote sensing data (NDVI) and multiscalar drought indices (SPI) already been considered by Vicente-Serrano et al. 2006 to model wheat and barley yields in Spain. Vicente-Serrano et al. 2006 have found that the inclusion of NDVI in a linear regression model based on SPI (February at 1-month time-scale) increases the model's performance. Moreover, Vicente-Serrano et al. 2006 have shown the potential of the combined use of NDVI and SPI to predict cereal production four months prior to harvest. Similarly, we also addressed the ability to estimate crop yield during growth stages early enough before harvesting. Table 2.1 indicates that the models based on remote sensing indices depend largely on the weekly values of mid-winter (December and January) and mid-spring to early-summer (late-April to June), suggesting a predictive power of crop yield based of the satellite-based data. The selection of SPEI of February, April, May and June in cluster 2 also suggests the predictive power of a range of drought time-scales for crop-loss estimation.

The MLR and ANN results suggest that the models displaying the strongest relationships are the same in both statistical techniques, and the strength of the statistical relationships found by the linear and non-linear methods is quite similar (Table 2.2 and Table 2.3). However, regarding the 4 models with the strongest relationships for the two cereals in the two cluster (W1a, B1b, W2c and B2c), the ANN techniques improve the MLR models except in the case of barley in cluster 2 (Table 2.3 and Figure 2.7). The explained variance of the model W1a using MLR increases 12% using ANN techniques, models B1b and W2c increase 7% and 2% respectively.

Despite the slightly overperforming of the ANN over the MLR techniques in 3 of the best 4 models (W1a, B1b, W2c and exception of B2c), the ability to estimate yield losses is overrated (HR and SR display higher values using ANN rather than MLR but FB values by ANN are generally indicative of overestimation). The cereal models in cluster 2 display the lower values of SR, HR and higher values of FAR, in comparison with the cluster 1, suggesting that despite the ability of SPEI in representing the average variability in the southern sector, it underperforms the estimation of crop-loss in comparison with remote sensing indices in cluster 1. However, most of the crop-loss events are estimated (high values of HR) by all models, suggesting the potential of the proposed methodology for the modelling of wheat- and barley-losses in IP.

A substantial number of studies have already suggested the better performing skills of ANN in comparison to MLR in cereal yield modelling (Jiang et al., 2004; Matsumura et al., 2015). In the Mediterranean region in particular, Incerti et al. (2007) proposed a drought risk analysis based on ANN for South Italy based on precipitation, temperature, evapotranspiration, NDVI and land cover. Climate change impact on durum wheat over the Mediterranean basin has been addressed by Ferrise et al. (2011) based on ANN as well. Using other alternative statistical techniques, such as partial least square regression, Hernandez-Barrera et al. (2017) had analysed the climate change impacts on wheat yield over Spain. Results by Ferrise et al. (2011) suggested that the projected warmer and drier climate will increase the risk of yield loss in the Mediterranean, and Hernandez-Barrera et al. (2017) suggested that climatic warming will lead to about 32% decrease in Spanish wheat production in the twenty-first century. Henceforth, an improved assessment of the agricultural crop yield impacts under current drought conditions is becoming crucial in a climate change context. The establishment of novel statistical techniques for crop modelling, such as ANN, constitutes an important step towards developing strategies to mitigate agricultural drought risk.

Besides some slight overestimation of yield losses, limitations of the presented results arise from the lack of forecasting of future yield losses of wheat and barley. Nevertheless, the present study indicates that based on mid-winter and mid-spring drought indicators, the estimation of the harvestable yield is predictable for the current year. In addition, the results from the calculation of the drought index SPEI using climate projections of precipitation and temperature, and further application using the statistical relationships found in the present study, would be rather interesting to compare with recent works. Other potential usefulness of this study for future research is to evaluate the suitability of the regional-scale crop yield models to each province of the IP individually. More future work should also cover other agro-areas of the IP and look towards the development of crop-specific agricultural drought risk models (e.g. using a probabilistic approach) based on the established models.

In summary, the statistical methodology used in this analysis relied on yield information at the province scale, and the results have shown the potential of crop yield modelling based on multiscalar (SPEI) and remote sensing (VCI, TCI and VHI) indices, using two empirical techniques (MLR and ANN), providing estimations of drought-impacts over large areas. In contrast, numerous modelling tools integrating the complex biophysical interactions of crop

growth (mechanistic crop simulation models) have been used by several authors (Paredes et al., 2014, 2016; Giménez et al., 2016), generally requiring careful calibration and several in-situ measurements, usually limited to the local/field scales. The model outcomes using the presented methodology are suitable for broader scales, and highlight the usefulness of such analysis in the framework of developing an agricultural drought risk model for cereal yields in the IP. In terms of an operational point of view, the results aim to contribute to an improved understanding of crop yield management under dry conditions, particularly regarding rainfed winter crops. Moreover, the present study will provide some guidance on user's decision-making process in agricultural practices in the IP, assisting farmers in deciding whether to purchase crop insurance.

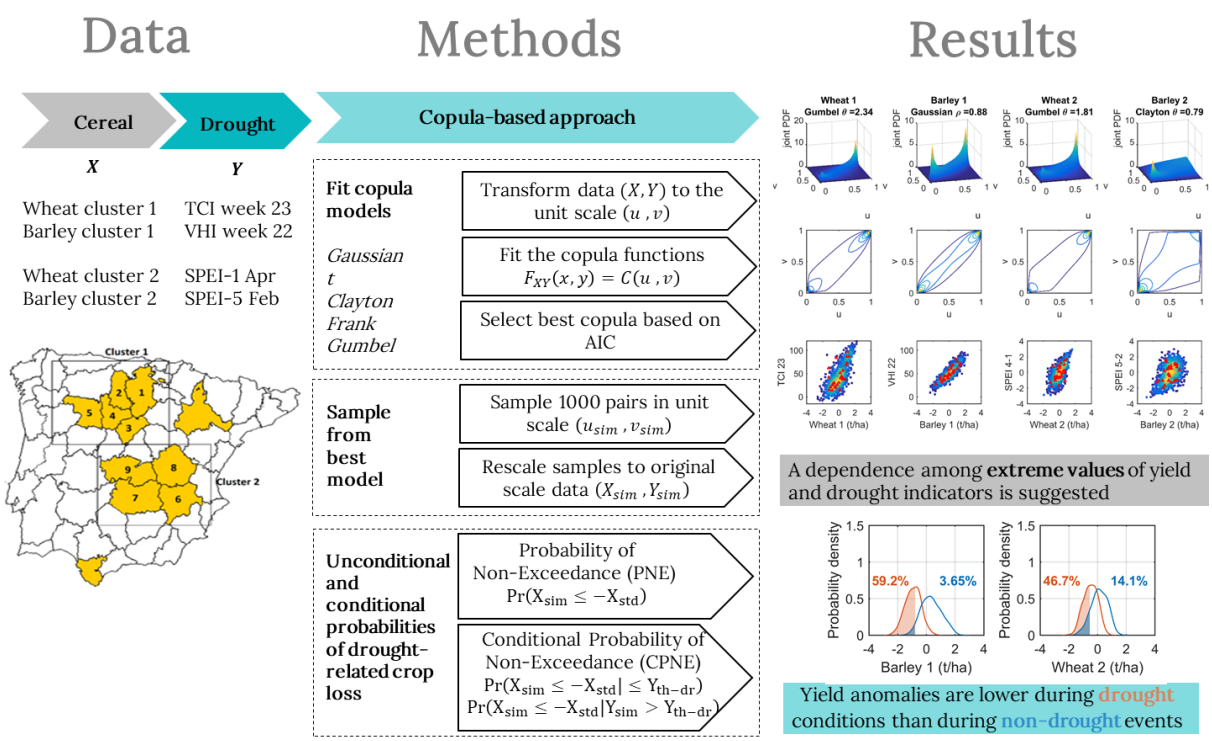
Chapter 3

Probabilistic modelling of the dependence between rainfed crops and drought hazard

***Ribeiro, A.F.S.**, Russo, A., Gouveia, C.M., Páscoa, P., Pires., C. A. L. (2019). Probabilistic modelling of the dependence between rainfed crops and drought hazard, Natural Hazards and Earth System Science, 19, 2795–2809, https://doi.org/10.5194/nhess-19-2795-2019*

***Andreia F. S. Ribeiro contributions:** preparation of the data (crop yields, remote sensing and SPEI); statistical analysis and interpretation of the results (copula-based analysis and estimation of conditional probabilities); production of all the figures and coordination of the manuscript writing.*

Graphical abstract



Highlights

- Copula functions are used to characterize the joint behavior of crop yields and drought conditions
- Copula functions generally point to a dependence among extreme values of yield based on asymmetrical structures
- The conditional probabilities of crop-loss indicate an increased risk of wheat- and barley-loss when there is a transition from non-drought to drought conditions
- The use of unconditional probabilities of crop-loss may lead to the underestimation of the agricultural impacts
- The northern cluster (cluster 1) seems to be more susceptible to drought-related barley-losses, while drought-related wheat-losses are more likely to occur in the southern cluster (cluster 2)
- The use of copulas for the analysis of the co-occurrence of dry and low-yield extreme events has proven to be a valuable methodology for the estimation of drought-related crop failure

Abstract

Extreme weather events, such as droughts, have been increasingly affecting the agricultural sector causing several socio-economic consequences. The growing economy requires improved assessments of drought-related impacts in agriculture, particularly under a climate that is getting drier and warmer. This work proposes a probabilistic model which intends to contribute to the agricultural drought risk management in rainfed cropping systems. Our methodology is based on a bivariate copula-approach using Elliptical and Archimedean copulas, which application is quite recent in agrometeorological studies. In this work we use copulas to model joint probability distributions describing the amount of dependence between drought conditions and crop yield anomalies. Afterwards, we use the established copula models to simulate pairs of yield anomalies and drought hazard, preserving their dependence structure, to further estimate the probability of crop-loss. In the first step, we analyse the probability of crop-loss without distinguishing the class of drought, and in a second step we compare the probability of crop-loss under drought and non-drought conditions. The results indicate that, in general, Archimedean copulas provide the best statistical fits of the joint probability distributions, suggesting a dependence among extreme values of rainfed cereal yield anomalies and drought indicators. Moreover, the estimated conditional probabilities suggest that when drought conditions are below the moderate thresholds, the risk of crop-loss increases between 32.53% (cluster 1) and 32.6% (cluster 2) in the case of wheat, and between 31.63% (cluster 2) and 55.55% (cluster 2) in the case of barley. From an operational point of view, the results aim to contribute to the decision-making process in agricultural practices.

Keywords: Drought-induced yield losses; Rainfed cereals; Copulas; Probabilistic decision making

3. Probabilistic modelling of the dependence between rainfed crops and drought hazard

3.1 Introduction

Agriculture is one of the activities most directly influenced by climate variability (Lesk et al., 2016; IPCC, 2012) and particularly by extreme weather events (IPCC 2012). The latter are a major source of risk in agricultural systems, often entailing substantial crop yield losses (Kogan et al., 2015a; Saadi et al., 2015; Bokusheva et al., 2016). Despite the constrains associated with the application of certain governmental policies in the agricultural systems, the success of mitigating the consequences of climate extremes is largely dependent on the development of appropriate risk management strategies (Quiroga et al., 2011; Paredes et al., 2014). For this purpose, probabilistic information of the agricultural risk associated to certain meteorological conditions is currently a major requirement (Iglesias and Quiroga, 2007; Jayanthi et al., 2014; Madadgar et al., 2017), particularly under the scope of the projected climate changes (Ferrise et al., 2011; Hernández-Barrera and Rodríguez-Puebla, 2017).

From both researcher's and stakeholder's perspective, the management of agricultural drought risk has been a challenging task for decades, mainly in regions dominated by high precipitation variability and recurrent dry and warm episodes, such as the Mediterranean region and in particular the Iberian Peninsula (IP) (Martin-Vide and Lopez-Bustins, 2006; Sousa et al., 2011; Vicente-Serrano et al., 2014). Recent works have found significant negative trends of drought indexes in the IP based on long-term time-series including the entire 20th century, particularly in southern regions (Sousa et al., 2011; Páscoa et al., 2017a), and the expected declining of crop yields due to future warming conditions is being pointed out (Ferrise et al., 2011; Hernández-Barrera and Rodríguez-Puebla, 2017).

The assessment of yield variability based on crop and meteorological information is crucial for a more stable farmer income and management (Reidsma et al., 2010). The recently developed drought index SPEI (Standardized Precipitation Evapotranspiration Index (Vicente-Serrano et al., 2010)) is found to be particularly suitable for agricultural drought applications in Mediterranean regions (Zampieri et al., 2017) and shows significant correlations with crop

yields in the IP (Páscoa et al., 2017b; Ribeiro et al., 2019b). On the other hand, crop models describing the biological processes are one of the existing tools used to assess crop productivity, e.g. CERES (Crop Environment REsource Synthesis) models (Hlavinka et al., 2010; Capa-Morocho et al., 2016a) and AquaCrop (Vergni et al., 2015; Paredes et al., 2016). These crop models are important tools in agrometeorological studies being able to compute irrigation requirements and yield simulations, and have been particularly useful for assessing the impacts of climate change on agricultural productions (Hlavinka et al., 2010). However, such models are limited in their ability to quantify the impact of climate variability on crop yields over larger scales (Estes et al., 2013) and the detailed representation of crop's biophysical interactions requires demanding parameterization settings and input data (Paredes et al., 2014, 2016; Giménez et al., 2016). Thus, empirical modelling constitutes an alternative to represent the large-scale impacts of drought conditions in the agricultural sector (Vicente-Serrano et al., 2006; Kogan et al., 2015a; Matsumura et al., 2015; Bokusheva et al., 2016) requiring lower computation costs than mechanistic modelling (Ferrise et al., 2011; Estes et al., 2013).

In addition, the use of satellite-based data is increasing for agricultural purposes (Rojas et al., 2011; Kogan et al., 2015a), and considerable correlations between remote-sensing of vegetation and crop yield are found in the IP (Vicente-Serrano et al., 2006; Gouveia and Trigo, 2008; Ribeiro et al., 2019b). Some studies have considered the use of different remote sensing drought indicators to account for different crop sensitivities to drought, such as to moisture and thermal conditions over the vegetative cycle (Kogan, 2001; Zarei et al., 2013; Bokusheva et al., 2016; Ribeiro et al., 2019b). Moreover, the establishment of models for estimating crop yield under drought influence, using the combination of different drought indicators and different time-scales of drought occurrence, have shown an added value in the performance of the crop yield simulations over the IP (Vicente-Serrano et al., 2006; Hernandez-Barrera et al., 2017; Ribeiro et al., 2019b).

The statistical modelling of crop yield variability under drought conditions has been previously done to estimate drought-related crop-losses (Kogan et al., 2015a; Zampieri et al., 2017; Ribeiro et al., 2019b). Some authors have estimated crop yield probability distribution functions to find crop-specific risk levels and have applied Monte Carlo methods to generate large sample sizes of yield distributions over Mediterranean areas (Iglesias and Quiroga, 2007; Resco et al., 2010). At the country-level in Europe, Naumann et al. (2015) have developed drought damage

functions using a single power-law dependence between drought severity and the associated damage. At a regional level in the IP, regression techniques (Hernandez-Barrera et al., 2017; Hernández-Barrera and Rodríguez-Puebla, 2017; Ribeiro et al., 2019b) and artificial neural network (ANN) models (Ribeiro et al., 2019b) have been used to model the response of rainfed winter cereal yields to drought conditions. A major conclusion in Ribeiro et al. (2019b) was that there are stronger relationships between remote sensing indices and cereal yield in the northern sector of the IP, and between SPEI and cereal yield in the southern sector of the IP. This character of the response of crop yields to climate conditions highlights how it varies according to the location, type of crop, moment of the vegetative cycle, drought indicator and temporal scale.

More recently, copula-based models have been applied for agricultural purposes, to model the dependence structures between crop yields and environmental conditions using joint distributions (Li et al., 2015; Bokusheva et al., 2016; Madadgar et al., 2017; Ribeiro et al., 2019a). The concept of copulas is quite popular in financial risk modelling, and recently is becoming a valuable tool to model the risks associated to climate hazards, such as droughts (Serinaldi et al., 2009; Ganguli and Reddy, 2012; Mirabbasi et al., 2012). Based on the Sklar's theorem (Sklar, 1959) a copula approach "joins" the probability of drought occurrence and the probability of crop-losses caused by the drought event. A detailed description about the use of copulas is provided by Nelsen (2013).

A major advantage of copula methods is the generation of joint distributions independently of their marginal distribution functions (Nelsen, 2006; Maity, 2018). Copula functions show a great flexibility in modelling the dependence between individual variables (such as crop yield and drought indicators) with complex relationships without making heavy assumptions. In addition, copula functions are adequate for modelling rare events in multivariate distributions and to generate large samples, allowing to find the probability that individual variables will not exceed a certain extreme (tailed) value (Madadgar et al., 2017). A recent study by Madadgar et al. (2017) have produced probability distributions of rainfed crop yields in Australia under drought impacts based on copula-based techniques, using the Standardized Precipitation Index (SPI) and the Standardized Soil-moisture Index (SSI). For crop insurance purposes at the farm-level in Kazakhstan, Bokusheva et al. (2015) modelled the joint distributions of wheat yields and two satellite-based drought indices (Vegetation Condition Index (VCI) and Temperature

Condition Index (TCI)). At the global scale, Leng and Hall (2019) have also used copulas to assess the likelihood of yield loss in response to droughts based on SPI for the a historical period (1961–2016) and future period (2071–2100) under the RCP8.5 emission scenario to investigate future changes in yield loss risk. The authors found that global wheat is more vulnerable to droughts than maize, rice and soybeans, and that global warming is expected to amplify drought-driven yield loss risk.

In this study, a copula-based approach is adopted to model the joint probability density function of crop yield and the drought conditions for probabilistic yield assessment, based on the data and empirical analysis previously considered in Ribeiro et al. (2019b). This method allows to estimate the dependence structures between the probability distributions of crop yield and drought indicators using copula functions. The novelty and interest of this approach relates to the fact that this methodology will allow to estimate the likelihood of crop-loss and compare the expected losses under drought conditions and non-drought conditions in the IP. This key question poses as a current demand with most interest to stakeholders, such as farmers and insurance companies, to mitigate agricultural drought risk over the major agricultural areas in the IP.

3.2 Data and Methods

3.2.1 Study area and data

The exposure analysis performed by (Ribeiro et al., 2019b) allowed the identification of two clusters of provinces in the IP dominated by rainfed agricultural practices (Figure 3.1), located approximately in the regions of Castilla-Léon (cluster 1 - northern region) and Castilla-La Mancha (cluster 2 – southern region). Given the suitability of using these two clusters for an agricultural drought analysis at the regional level, here we have considered the same area selection criteria: provinces with more than 50% of the territory occupied by agricultural areas and more than 50% of rainfed crops according to the CORINE Land Cover (2012) (for more details please see Ribeiro et al., (2019b)). Considering previous requirements and for sequential purposes, the crop and drought hazard data used in (Ribeiro et al., 2019b) have been incorporated in the present study to analyse the distributions of probabilities. Spatial averages

of annual yield anomalies (t/ha) of barley and wheat were computed over the two clusters during the period of 1986-2012, based on production (tons, t) and area (ha) information obtained from the Portuguese National Statistics Institute and the Spanish Agriculture, Food and Environment Ministry.

Drought conditions were investigated using the hydro-meteorological drought indicator SPEI and three satellite-based indices obtained from NOAA-AVHRR since 1981, namely the VCI (Kogan 1990), the TCI (Kogan 1995) and the Vegetation Health Index (VHI) (Kogan 1995). The monthly drought index SPEI gridded values, with spatial resolution of 0.5°, were computed based on precipitation and temperature values from the Climate Research Unit TS3.21 database (Harris et al., 2014), using a variety of time scales (1 to 12 months). The weekly global maps of VCI, TCI, and VHI were retrieved at 4km spatial resolution from NOAA's ftp server (ftp://ftp.star.nesdis.noaa.gov/pub/corp/scsb/wguo/data/VHP_4km/geo_TIFF/). While SPEI computation uses climatic water balance anomalies incorporating the role played by the evaporative demand on the occurrence of dry events (Vicente-Serrano et al., 2010), the remote sensing indices characterize the moisture, through the VCI, the temperature induced stress through the TCI and health of vegetation, through the VHI.

Considering the vegetative cycle of wheat and barley, and in accordance with the results obtained by Ribeiro et al. (2019b), the data of VCI, TCI, and VHI used in this work covered the period from week 35 (early September) to week 25 (late June), and data of SPEI covered January to June. Spatial averages of all these indicators were computed for each provincial cluster and used for further modelling of the joint probability between the drought hazard and cereal yield anomalies over the period 1986-2012. Stepwise regression models (95% confidence level) were established to select the time scales and months of SPEI, together with the weeks of VCI, TCI, and VHI better related with wheat and barley annual yield (Ribeiro et al., 2019b). The selection of the most relevant drought indicator for each cereal and cluster was performed based on the largest absolute value of the standardized regression coefficients from the models developed in Ribeiro et al. (2019b), in order to constitute pairs of cereal yield anomalies and drought indicators. Afterwards, for each cereal time series, the joint probability with drought conditions was estimated using one drought indicator.

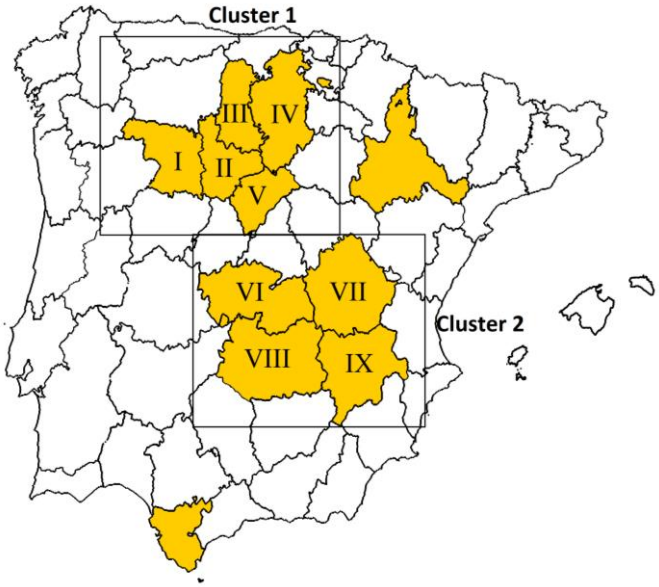


Figure 3.1 - Provinces with more than 50% of the territory occupied by agricultural areas and more than 50% of rainfed crops (yellow) according to CLC2012, and selected clusters of provinces. Cluster 1 provinces: Zamora (I), Valladolid (II), Palencia (III), Burgos (IV) and Segovia (V). Cluster 2 provinces: Toledo (VI), Cuenca (VII), Ciudad Real (VIII) and Albacete (IX).

3.2.2 Copula-based method

3.2.2.1 The concept of copula

Copula functions are powerful tools used to estimate the joint distribution between variables (Zhang et al., 2011a; Bokusheva et al., 2016; Madadgar et al., 2017). The concept of copula was firstly introduced by Sklar (1959) to decompose a joint cumulative distribution function $F_{XY}(x, y)$ into two parts (Equation 3.1): the marginal distribution functions $F_X(x) = u$ and $F_Y(y) = v$, and the copula *C* describing the dependence part of the joint distribution,

$$F_{XY}(x, y) = C(u\,, v), \tag{3.1}$$

where the margins u and v are uniformly distributed on the interval [0,1] (Nelsen, 2006). This study adopts a bivariate modelling approach such that, for each pair (X, Y) of cereal and drought indicator over each cluster we considered bivariate copula functions to estimate the joint probability distributions. Trivariate copulas have been proposed in the analysis of hydrological extremes (Bezak and Brilly, 2014; Saghafian and Mehdikhani, 2014; Afshar et al., 2016), but the development of higher dimensional copulas exhibits very complex structures and further studies and evaluations are required. In comparison to high-dimensional copulas, the two-

dimensional copulas involve much less computational cost and allows for more easily interpretable and illustratable relationships between the interval margins. For this reason, in the present study we restricted the analysis to the bivariate case using two-dimensional copulas, simplifying the results interpretation.

There is a range of copula families described in the literature which are able to estimate the dependence between the univariate variables (Nelsen, 2006). The most commonly used copula families focus on the Archimedean and Elliptical classes (Maity, 2018). There are three Archimedean copulas particular popular given their simple functional form and their different patterns of dependence captures – Clayton, Gumbel and Frank – while there are two most popular Elliptical copulas derived from Elliptical distributions – Gaussian and t-copula. These five copula functions are well-documented and have been employed in recent agrometeorological studies with a number of annual observations similar to our study (Bokusheva et al., 2016; Madadgar et al., 2017; Zscheischler et al., 2017). Table 3.1 summarizes the mathematical expressions of the referred copula functions considered in the present study.

Table 3.1 - Equations of the copula functions, where u and v are univariate variables, Φ^{-1} is the inverse of standard Gaussian CDF, ${t_{df}}^{-1}$ is the inverse t–student CDF, df is the degree of freedom, ρ and θ are dependence parameters.

Family	**Joint Cumulative Distribution Function** $C(u, v)$	**Parameter range**
Gaussian	$\int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} e\left(-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right) dudv$	$-1 \leq \rho \leq 1$
t	$\int_{-\infty}^{{t_{df}}^{-1}(u)} \int_{-\infty}^{{t_{df}}^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} e\left(1+\frac{u^2+v^2-2\rho uv}{df(1-\rho^2)}\right)^{-\frac{df+2}{2}} dudv$	$-1 \leq \rho \leq 1$ $df \geq 1$
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}}$	$\in [-1, \infty[/ \{0\}$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$\theta \neq 0$
Gumbel	$e^{-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{\frac{1}{\theta}}}$	$\lvert\theta\rvert < \infty$

An important concept for studying extreme events is the tail dependence, whose importance is more critical than the overall dependence structure for risk analysis (Bokusheva, 2014). The joint tail behaviour describes the amount of dependence in the corners of upper-right and lower-

left quadrants (i.e. joint extreme events), and its representation depends on the type of copula (Nelsen, 2013). The Frank, Gaussian and t-copula describe a joint symmetric structure with a symmetric tail dependence, i.e. the same degree of dependence in both pairs of extremes. The Clayton/Gumbel copulas have an asymmetric tail dependence with greater dependence in the lower/upper tail suggesting greater probabilities of joint lower/upper extremes (i.e., lower/higher values of yield anomalies given lower/higher values of drought indicators).

3.2.2.2 Fitting of the copula functions

The estimation of the copula parameters can be performed using different methods based on maximum likelihood, such as Maximum Likelihood Estimate (MLE), Inference Functions for Margins (IFM) or Canonical Maximum Likelihood (CML) (Maity, 2018). With MLE, both individual margins and copula parameters are estimated together, whereas with IFM the marginal parameters are first estimated individually. In this study the statistical inference of the copula functions is performed with the CML method, which stands for a nonparametric estimation of the margins. In this way, the individual variables were first transformed to the unit scale (pseudo-observations) using the kernel density estimator of the Cumulated Distribution Function (CDF), without making assumptions about the marginal distributions (Figure 3.2). The drawback of the shorter sample size is surpassed by the nonparametric estimation of the margins, which avoids heavy assumptions about their distributions, even when the available sample is rather small (Fahr, 2017; Corder and Foreman, 2011). The fitting of the bivariate copula functions was then applied to the pseudo-observations and the dependence parameters were estimated by means of maximum likelihood (Figure 3.2). Figure 3.2 summarizes the main steps of the copula-based approach adopted in the present study. For a detailed description on fitting methods please see Maity (2018).

The Akaike's Information Criteria (AIC) is frequently employed as a model selection tool in copula modelling (Mirabbasi et al., 2012; Li et al., 2015). Therefore, the selection of the best copula function for each pair of cereal and drought indicator was made based on the evaluation of AIC values calculated as AIC = –2 × (sum of log-likelihood) + 2 × (number of parameters) (Figure 3.2). The copula function minimizing the AIC value was selected for each case. For verification purposes, the leave-one-out cross-validated log-likelihood was also computed

during the estimation of the parameters. This step was performed to confirm the reliability of the selected copula models and we found that, in general, the same functions are selected with both the AIC and the cross-validated log-likelihood criteria. For this reason, and given the wide use of the AIC, only the results for model selection based on the AIC will be presented.

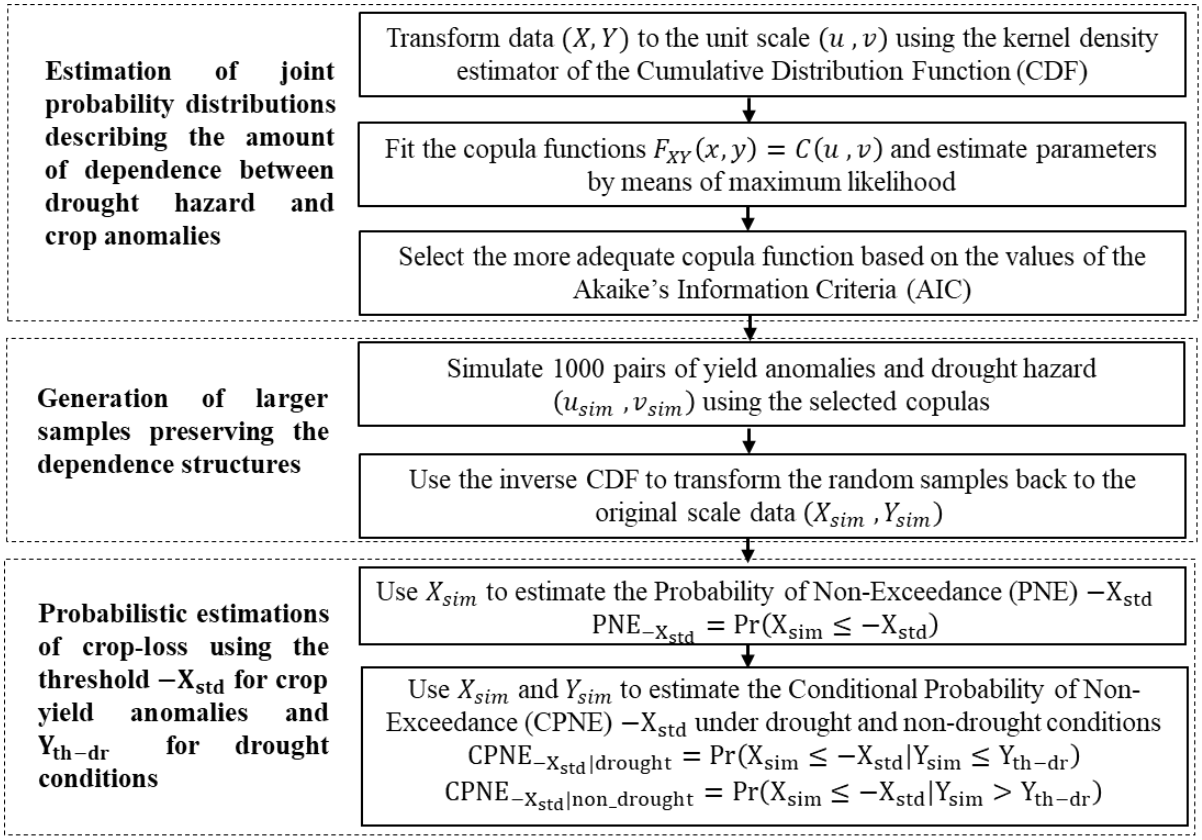


Figure 3.2 - Scheme of the copula-based approach adopted in the present study.

3.2.2.3 Probability of Non-Exceedance and Conditional Probability of Non-Exceedance

After the estimation of the copula parameters, the established models are used to simulate 1000 pairs of uniformly distributed data (Figure 3.2). In the present study, let $F_{X_{sim}}(x) = u_{sim}$ denote the simulated CDF of yield x and $F_{Y_{sim}}(y) = v_{sim}$ the simulated CDF of drought indicator y. The data generation using the joint relationship preserves the dependence structure between the margins. The simulated data in the range [0, 1] is transformed back to the original scale using the kernel estimations of the inverse CDF, originating X_{sim} and Y_{sim}, respectively.

First the copula simulations were used to estimate the risk of crop-loss in terms of the probability of not exceeding a threshold value of yield, i.e., Probability of Non-Exceedance (PNE) (Figure 3.2). In this study we considered the threshold of minus one standard deviation ($-X_{std}$) of each cereal yield anomalies time-series, as we are focused in real losses of yield and not just values below the mean (Equation 3.2).

$$PNE_{-X_{std}} = u_{sim}(-X_{std}) = Pr(X_{sim} \leq -X_{std}) \tag{3.2}$$

The PNE gives information about how likely the occurrence of a yield value below a certain threshold is. In other words, it gives the expected chance in percentage that the negative yield anomaly will not exceed (i.e. is not higher than) minus one standard deviation (-1 std).

Afterwards we have partitioned the simulated data points of X_{sim} into those corresponding to drought (e.g. SPEI <= -0.84 (Agnew, 2000) and/or VHI <= 40 (Kogan, 2001)) and non-drought conditions (e.g. SPEI > -0.84 and/or VHI > 40) (Figure 3.2). The respective CDFs were used to estimate the risk of crop-loss in terms of the Conditional Probability of Non-Exceedance (CPNE) given by Equation 3.3 and Equation 3.4, where Y_{th-dr} is the drought threshold, respectively -0.84 and 40 when the SPEI and VHI/TCI are used.

$$CPNE_{-X_{std}|drought} = Pr(X_{sim} \leq -X_{std} | Y_{sim} \leq Y_{th-dr}) \tag{3.3}$$

$$CPNE_{-X_{std}|non\_drought} = Pr(X_{sim} \leq -X_{std} | Y_{sim} > Y_{th-dr}) \tag{3.4}$$

For the purpose of validation and estimation of confidence intervals, the theoretical values of the above CPNE were inferred from the copula functions using the Equation 3.5 and Equation 3.6 (deduced from the definition of conditional probability),

$$CPNE_{-X_{std}|drought} = \frac{C(u_{-std}, v_{th-dr})}{v_{th-dr}} \tag{3.5}$$

$$CPNE_{-X_{std}|non\_drought} = \frac{u_{-std} - C(u_{-std}, v_{th-dr})}{1 - v_{th-dr}} \tag{3.6}$$

where $u_{-std} = F_X(-X_{std})$ and $v_{th-dr} = F_Y(Y_{th-dr})$ are the marginal probabilities of crop-loss and drought occurrence obtained from the kernel-based univariate CDFs. The lower and upper bound of the 95% confidence interval (ci) of the estimated copula dependence parameters were

considered using the Equation 3.5 and Equation 3.6 in order to obtain the confidence interval of CPNE coming from the inaccuracy of the copula parameter and to address if the CPNE using simulations (Equation 3.3 and Equation 3.4) lies within the 95% confidence level.

In sum, first we describe the joint probability of drought hazard and yield anomalies and simulate pairs of data preserving their dependence structure. After that, probability of crop-loss (PNE) and conditional probability of crop-loss (CPNE) are estimated, addressing whether the probability of crop-loss under drought conditions is higher than during non-drought conditions, and if distinguish drought severity is important. The probability distributions (based on a normal kernel function) of the generated yield anomalies are also analysed for graphical visualisation of the area corresponding to crop-loss.

3.3 Results

3.3.1 Fitting copula models

The estimates of the dependence between the yield anomalies and drought hazard were performed using the selected drought indicators resumed in Table 3.2. This selection of drought indicators highlights that the response of crop yields to climate conditions vary according to the location, type of crop, moment of the vegetative cycle and the temporal scale. While annual yield anomalies in cluster 1 are better characterized by short-term responses to the drought conditions based on the weekly values of TCI and VHI, the annual yield anomalies in cluster 2 are better characterized by the monthly response to the dry conditions based on the SPEI. In terms of predictability, the effects of temperature (TCI) and vegetation health (VHI) during late growth stages (weeks 23 and 22 correspond approximately to beginning of June and end of May, respectively for wheat and barley) are the most influential conditions in the northern cluster. On the other hand, the yields in cluster 2 are influenced by drought conditions described by SPEI much earlier, in the beginning of the intermediate growth stages (February and April with 5 and 1 month of time-scale, respectively for wheat and barley). In this way, the importance of including multiple drought response time scales is evidenced for predictability purposes and assessment of drought-related crop-losses.

Table 3.2 - Variables used for copula application. In the first column, the numbers 1 and 2 correspond to the respective provincial cluster (clusters 1 and 2). In the second column, the numbers correspond to the selected weeks in the case of the remote sensing indices, and to the selected months and time-scales (in months) in the case of SPEI. The values of the standardized regression coefficients were determined by Ribeiro et al. (2019b).

Cereal (X)	**Drought indicator** (Y)	**Standardized regression coefficients**
Wheat 1	TCI 23	0.76
Barley 1	VHI 22	0.91
Wheat 2	SPEI 4-1	1.05
Barley 2	SPEI 2-5	1.07

Figure 3.3 shows the non-parametric estimations of the CDF of the individual variables from Table 3.2, here used to transform the variables to the unit scale (pseudo-observations) for the copula modelling. A good agreement with the ECDF is suggested (Figure 3.3) and the crop-loss and drought thresholds used in this study ($-X_{std}$ and Y_{th-dr} respectively) are illustrated. A straightforward way of visualization of the association between the cereal yields and drought conditions was first carried out based on the scattering of the uniform pseudo-observations of the margins (Figure 3.3 – bottom). Most of the transformed data points are concentrated along the diagonal line (Figure 3.3 – bottom), mainly due to the correlations between the yield and selected drought indicators (Ribeiro et al., 2019b). Most of the works based on copulas have estimates of the marginal distribution functions (Mirabbasi et al., 2012; Afshar et al., 2016; Bokusheva et al., 2016), whereas this procedure has no requirement for prior knowledge of the marginal distributions, entailing therefore less heavy assumptions.

The estimates of the dependence between the yield anomalies and drought indicators were performed using the copula functions from Table 3.1 (Gaussian, t-copula, Clayton, Frank and Gumbel). Table 3.3 indicates each copula dependence parameter estimates (ρ, df or θ) and respective AIC values. Based on the values of AIC, a Gaussian copula, a Clayton copula and two Gumbel copulas were eligible to perform the best fits (Table 3.3). In general, the Archimedean copulas are better suited to estimate the joint distributions between crop yield and drought indicators in most of the cases (Table 3.3), with the exception of barley in cluster 1, which is better fitted by a Gaussian copula. Given that AIC penalizes the number of estimated parameters (Wilks, 2006), t-copulas are not expected to be chosen, since they have two parameters that control the tail dependence.

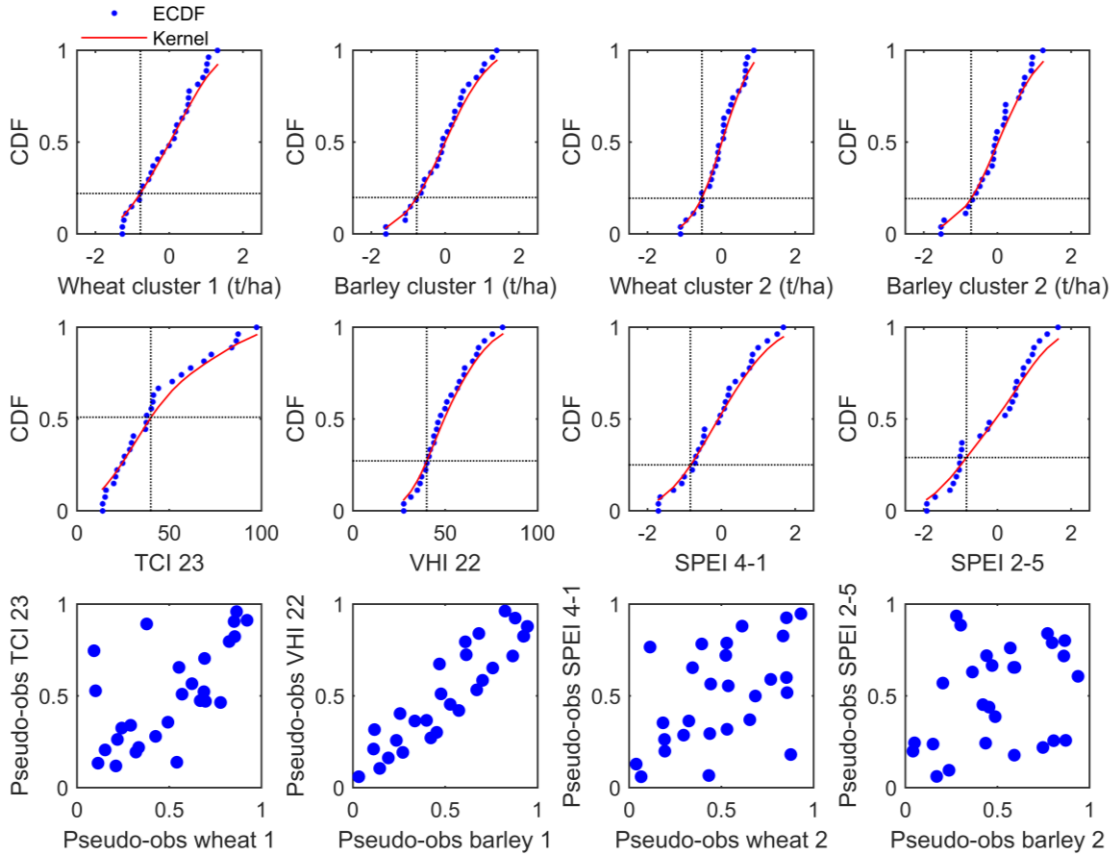


Figure 3.3 - Empirical cumulative distribution functions (ECDF, blue points), kernel density estimation of the CDF (red line), crop-loss and drought thresholds (dotted black vertical line), respective marginal probabilities of crop-loss and drought occurrence (dotted black horizontal line) and pseudo-observations (scatter) of the margins on the interval [0,1].

Table 3.3 - Copula dependence parameter estimates (ρ,df or θ), 95% confidence interval (ci) in parenthesis and AIC values. The ci 95% denoted by '-' indicates that the model was unable to compute the ci using the profile likelihood of the parameter. The selected models according to the lowest value of AIC are in bold.

	Gaussian			**t-copula**			**Clayton**			**Frank**			**Gumbel**		
	ρ	**ci 95%**	**AIC**	**df**	**ci 95%**	**AIC**	**θ**	**ci 95%**	**AIC**	**θ**	**ci 95%**	**AIC**	**θ**	**ci 95%**	**AIC**
W1	0.63	0.33;0.82	-11.79	0.75	0.02;2.23	-3.73	1.91	1.14,2.68	-11.07	6.45	3.95,8.95	-13.42	**2.34**	**1.72,2.96**	**-16.7**
B1	**0.88**	**0.80;0.96**	**-39.1**	0.92	-	4	4.09	2.21,5.96	-36.97	12.5	4.90,20.1	-38.93	3.11	1.89,4.33	-32.43
W2	0.54	0.15;0.74	-7.23	0.54	-0.25;2.82	-3.55	1.35	0.56,2.13	-7.95	4.35	1.83,6.88	-6.69	**1.81**	**1.24,2.38**	**-8.78**
B2	0.32	-0.07;0.62	-0.99	0.42	-14.02;21.96	3.02	**0.79**	**-0.38,1.95**	**-2.70**	2.54	-0.06,5.14	-1.06	1.42	0.96,1.88	-1.12

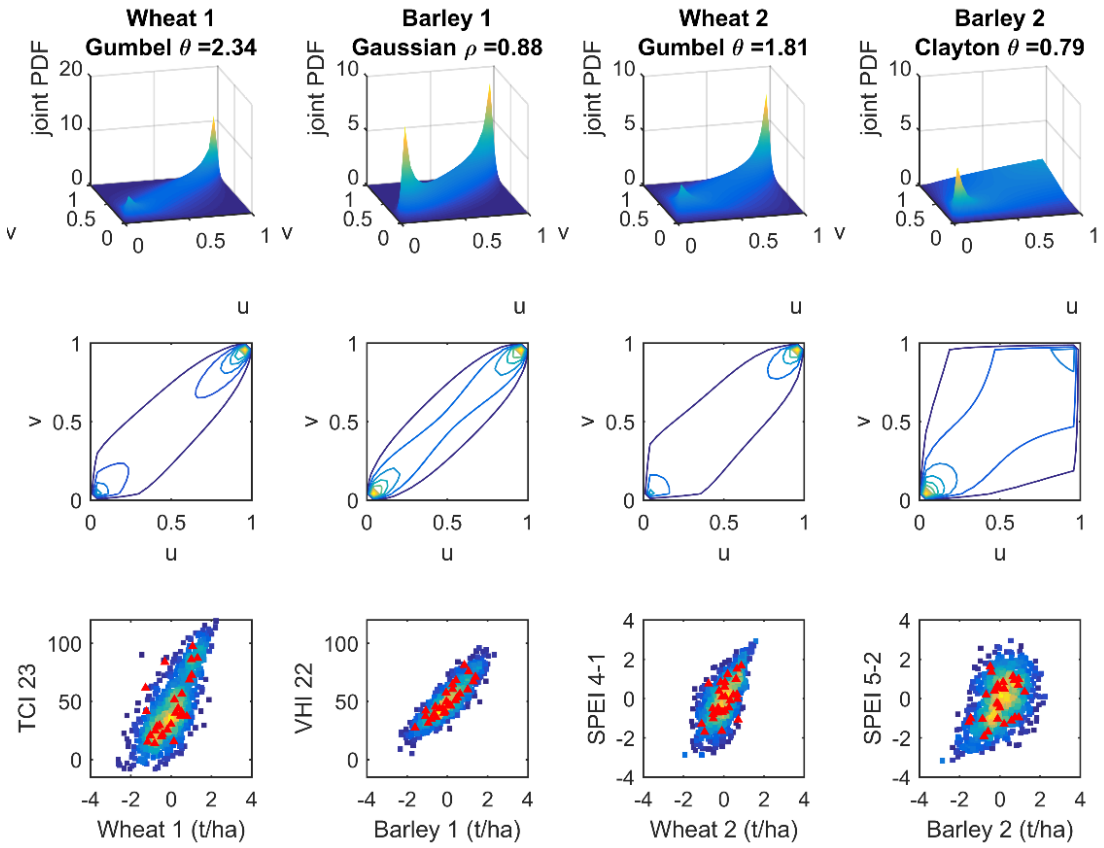


Figure 3.4 - Selected joint Probability Distribution Functions (PDF) where u and v are scalar values on the interval [0,1] (top), contours showing the two-dimensional view of PDFs (middle) and observed (red triangles) and copula-based simulations (density squares) scatter plots of crop yields and drought indicators (bottom).

The selected copula functions (Table 3.3) suggest that, in general, the relationship between yield and drought conditions is described by an asymmetric dependence in the tails of the joint distributions, except in the case of barley in cluster 1. This feature is illustrated in Figure 3.4, showing the different shapes and contours of the selected copula densities. While wheat in cluster 1 and 2 shows a stronger dependence in the upper tail of the joint distributions based on Gumbel copulas (suggesting higher probability of observing a higher value of yield anomalies given a high value of the drought indicators), barley in cluster 2 shows stronger dependence in the lower-left tail based on a Clayton copula, suggesting higher probability of finding a lower value of yield anomalies given a low value of the drought indicators. The randomly generated yield and drought data was transformed back to the original scales (Figure 3.4 bottom panel) and the respective scatter plots indicate that more extreme values are generated using the joint distribution relationships. In general, the modelling of the joint distributions leads to results close to the real observations (Figure 3.4 bottom panel).

3.3.2 Probability of Non-Exceedance and Conditional Probability of Non-Exceedance using copula simulations

After estimating the joint distribution functions and simulating pairs of data preserving the modelled dependence structures, we evaluate and compare the Probability of Non-Exceedance (PNE) and Conditional Probability of Non-Exceedance (CPNE) as a function of the crop-loss threshold. In this way, we evaluate if distinguishing drought severity leads to different risk values of crop-loss in comparison to disregarding a drought threshold (using only simulations of yield) and compare the probability of crop-loss under drought and non-drought conditions (by means of both simulations of yield and respective drought indicator). One of the key advantages of estimating the values of PNE and CPNE by means of the copula simulations is the use of larger samples which entail more joint extreme values based on the joint behaviour of crop yields and drought hazard.

Figure 3.5 shows the PNE curves and the distributions of the simulations of yield anomalies with the respective crop-loss area correspondent to the probability (%) of the yield anomaly not exceeding -1std. The PNE curves indicate more than 19% chance of having crop-losses in all cases. According to Figure 3.5, wheat at cluster 1 is the cereal with the highest risk level (22%) followed by barley in cluster 1 (19.8%), wheat in cluster 2 (19.4%) and barley in cluster 2 (19.2%) (Figure 3.5). As mentioned before, the wheat's left tail area (negative yield anomalies) is slightly higher in cluster 1, suggesting a higher risk of wheat-loss in the northern sector of the IP.

The following target was to compare the likelihood of crop-loss under drought and non-drought conditions. Figure 3.6 shows the simulated crop yield anomalies during drought (orange left-sided boxplots) and non-drought (blue right-sided boxplots) events. As expected, the boxplots show lower (and negative in average) values of yield anomalies during drought events in comparison with non-drought episodes. Although the number of samples simulated under drought conditions is smaller than under non-drought conditions (Figure 3.6), the use of copula simulations enhances the amount of simulated joint low extremes (i.e. co-occurrence of crop-loss and drought events).

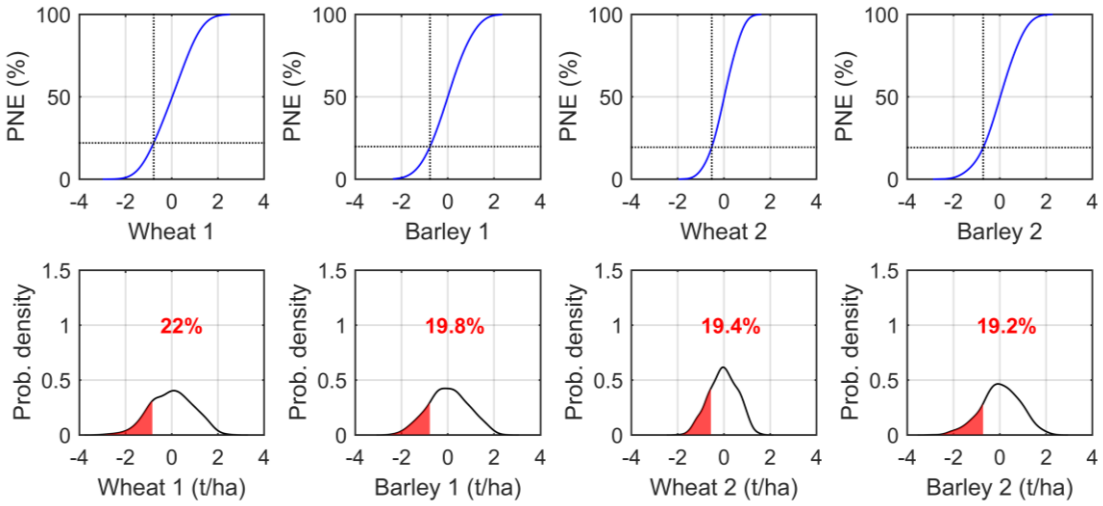


Figure 3.5 - Probability of non-exceedance (PNE) function (%) of yield anomalies (top) in both clusters based on the derived simulations from the estimated copulas and respective probability density estimates (bottom). On the bottom panels, the red values indicate the probability of crop-loss which is also indicated in the top panels by the intersected dashed lines indicating the threshold of crop-loss and respective PNE value.

The differences in terms of crop-losses between cereals and regions is much evident when differentiating the climatic conditions (Figure 3.7), particularly during drought conditions. Figure 3.7 shows that the values of CPNE under drought (non-drought) conditions are above (below) the values of PNE illustrated in Figure 3.5. In comparison with the distributions of yield simulations without conditioning to specific thresholds of the drought indicators shown in Figure 3.5 and Figure 3.7 the distributions of the yield simulations during drought events show a shift to the left towards negative values of yield anomalies, while the distributions of yield simulations during non-drought events show a shift to the right towards positive values of yield anomalies (Figure 3.7). The case of barley in cluster 1 is quite distinct exhibiting a drought-related barley-loss almost 3 times higher than the value illustrated in Figure 3.5 (19.8%), supporting the importance of conditional probabilities for agricultural drought risk purposes. The conditional probability of wheat-loss (Figure 3.7), is also higher when focusing on drought conditions, although it is less than two times the values shown in Figure 3.5.

Regarding the drought-related barley-loss, the distribution of barley in cluster 1 is more shifted to negative yield anomalies, stressing that the drought risk of barley-loss is higher on cluster 1 (59.2%) than on cluster 2 (39.4%), while it is quite similar on both clusters in Figure 3.5. While barley suggests higher conditional probabilities of crop-loss under drought conditions in cluster 1, wheat suggests higher conditional probabilities of crop-loss under drought conditions in

cluster 2 (46.7%) in comparison to cluster 1 (36.5%). Among all the cases, the highest level of drought-related crop-loss is 59.2% observed in the case of barley in cluster 1, followed by wheat in cluster 2 with 46.7% chances of crop-loss under dry conditions.

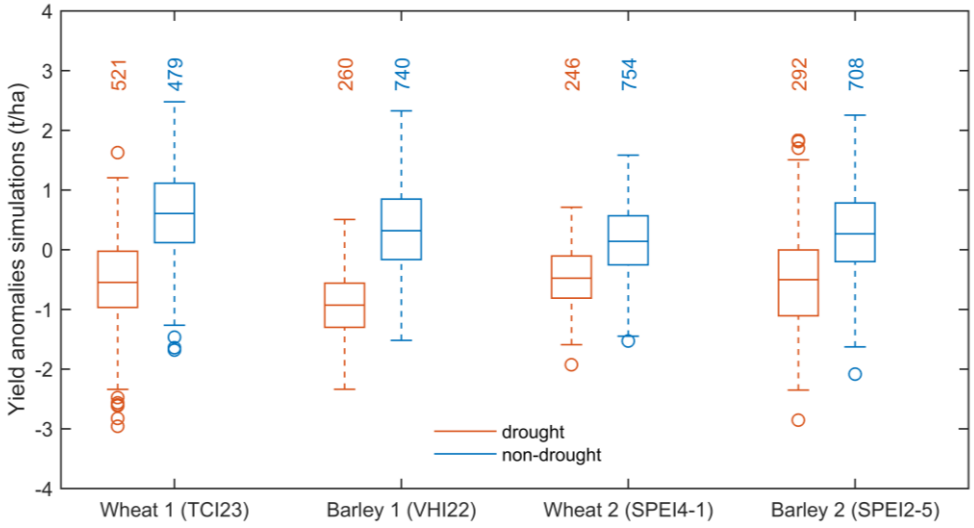


Figure 3.6 - Wheat and barley yield simulations differentiating drought (orange) and non-drought conditions (blue) according to the respective drought indicator denoted in parenthesis in the x-tick label. The numbers on top of the boxplots denote the sample size of the simulations under the different climatic conditions.

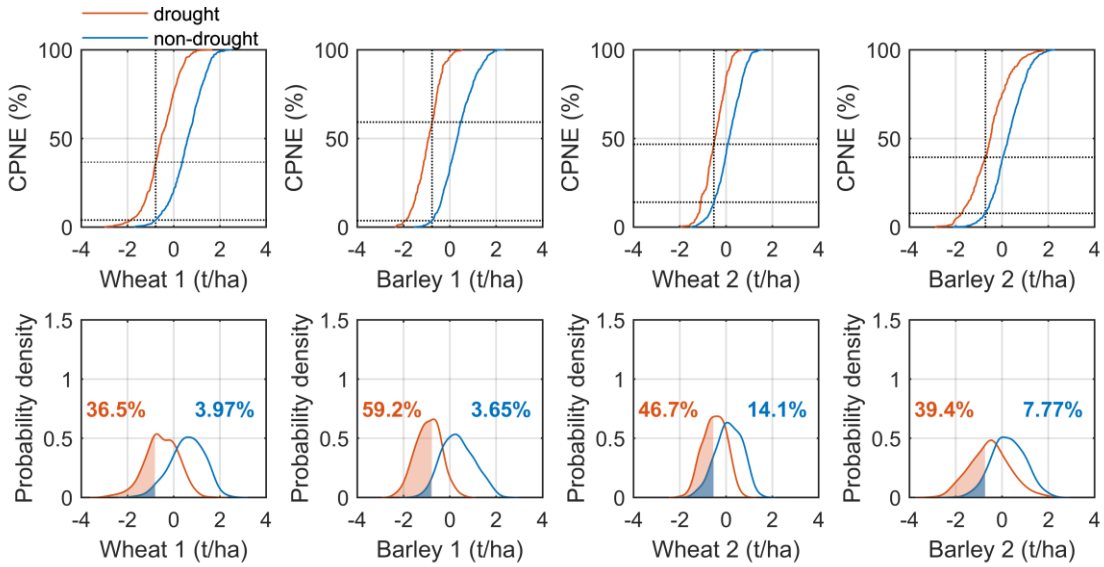


Figure 3.7 - Conditional probability of non-exceedance (CPNE) function (%) based on the derived copula simulations (top) and respective probability density estimates (bottom) under drought (orange) and non-drought conditions (blue). On the bottom panels, the orange and blue values indicate the probability of crop-loss under the different climatic conditions, which is also indicated in the top panels by the intersected dashed lines indicating the threshold of crop-loss and respective CPNE value.

The theoretical CPNE based on Equation 3.5 and Equation 3.6 (Table 3.4) agrees quite well with the estimates of the CPNE in Figure 3.7, thus corroborating the representativeness of the copula experiment using 1000 simulations. Nevertheless, the use of simulations allows to increase the sample size and to generate more joint extreme values based on the dependence structures characterized by the selected copulas. In addition, the effect of the copula parameters (ρ or θ) inaccuracy due to the finiteness of available sample is considered in Table 3.4 in terms of the 95% confidence level interval of CPNE based on the confidence interval of the copula parameters taken from Table 3.3 and Table 3.4 shows that the theoretical CPNE under drought conditions still remain well above the CPNE under non-drought, with their difference taking the smallest value at the lower bound of the copula parameter confidence interval. In most cases, those differences are positive, as expected from the effect of drought on crop yield, despite the relative finiteness of the sample to fit the copula models.

Table 3.4 - Theoretical CPNE (%) during drought and non-drought conditions (Equation 3.5 and Equation 3.6) and respective lower and upper bounds of the 95% confidence interval, where u_{-std} and v_{th-dr} are the marginal probabilities of crop-loss and drought occurrence, and θ or ρ are the estimated copula parameters with 95% confidence limits (Table 3.3). The only exception which gives greater values of CPNE during non-drought conditions rather than drought is denoted by '*'.

				Lower confidence bound (95%)			Upper confidence bound (95%)					
	Copula	u_{-std}	v_{th-dr}	**θ or ρ**	**CPNE drought**	**CPNE non-drought**	**θ or ρ**	**CPNE drought**	**CPNE non-drought**	**θ or ρ**	**CPNE drought**	**CPNE non-drought**
W1	Gumbel	0.22	0.51	2.34	39.3	4.00	1.72	35.4	8.44	2.96	41.2	1.98
B1	Gaussian	0.20	0.27	0.88	62.4	4.31	0.80	56.3	6.54	0.96	70.3	1.40
W2	Gumbel	0.19	0.25	1.81	42.5	11.2	1.24	27.8	16.1	2.38	51.5	8.17
B2	Clayton	0.19	0.29	0.79	41.1	9.99	-0.38	6.55*	24.1*	1.95	55.2	4.23

The results show that CPNE based on simulations (Figure 3.7) and theoretical equations (Table 3.4) indicate that the probabilities of crop-loss increase when drought conditions occur, even considering the two-sided confidence bound values of the copula parameters. Moreover, the results indicate that the CPNE using the simulations (Figure 3.7) lies within the estimates of CPNE using the two-sided confidence bound values of the copula parameter at the 95% level of confidence (Table 3.4). The only exception is the case of barley in cluster 2 considering the lower bound of θ, which gives greater probabilities of crop-loss during non-drought conditions rather than during drought conditions, suggesting that other factor than water stress is the cause of crop-failure. This result has to do with the negative value of the copula parameter in the lower confidence bound (θ = -0.38), thus suggesting a weak dependence between crop-loss and drought conditions in this case. However, at the 80% confidence level ($\theta \in [0.03, 1.55]$) the

values of the copula parameter confidence bounds are both positive and give higher CPNE under drought conditions. This lack of accuracy of the CPNE at the 95% in the case of barley in cluster 2 may be the reason why the CPNE under drought conditions are not the highest of all cases, as would be expected from a Clayton copula (which is known for capturing lower tail dependence).

3.4 Discussion

This study investigated the usefulness of copula methods in estimating the likelihood of drought risk in wheat and barley cropping systems, when applied to two regions in the IP. Here we proposed to model the joint probability of yield and drought hazard using copulas, based on a prior analysis of the association between drought and crop-loss (Ribeiro et al., 2019b). The advantage of using a probabilistic approach is to meet the ambitious challenge of helping farmers and stakeholders in managing their operations, by identifying the probability of crop-loss under specific drought conditions. Hernández-Barrera and Rodríguez-Puebla (2017) and Ferrise et al. (2011) have shown that projected warmer and drier climate will lead to wheat yield shortfall over the IP and Mediterranean, respectively, highlighting the importance of establishing novel statistical approaches for agricultural drought risk analysis. Other crops rather than rainfed cereals are also expecting significant losses during the next century in the IP (Saadi et al., 2015; Resco et al., 2010; Quiroga and Iglesias, 2009), and the here proposed crop-specific approach could be applied to other agricultural systems under drought conditions for different regions around the world.

The novelty of the presented models, in comparison to other works addressing climate risk in the IP (Iglesias and Quiroga, 2007; Resco et al., 2010; Ribeiro et al., 2019b), is the focus on the impacts associated with droughts and on the joint probability of rainfed yield anomalies and drought hazard. Previous works using copulas in hydro-climatology studies have tended to focus on the joint distribution of different characteristics of the hazardous events, such as frequency, intensity, severity, duration, among others (Li et al., 2015; Chen et al., 2013; Mirabbasi et al., 2012). Moreover, the restriction to the bivariate case allowed for a simpler interpretation of the results, in contrast to higher dimension copulas (Ganguli and Reddy, 2013;

Afshar et al., 2016), for instance by adding as copula variables, other factors influencing crop yield beyond drought.

More recently, copulas have been applied to estimate the joint behaviour of drought conditions and the associated impacts in agricultural systems (Bokusheva et al., 2016; Madadgar et al., 2017; Leng and Hall, 2019; Ribeiro et al., 2019a), instead of using drought information only. We have adopted a similar approach to reproduce time-, regional-, and crop-specific dependence of drought conditions, and the probability distribution of crop yield anomalies under drought conditions was estimated for risk analysis. In addition, the use of different drought indicators in this study represents an advantage since crops react differently to several factors at distinct moments and locations, highlighting the importance of quantifying the contributions of different drought indices on a regional scale (Zarei et al., 2013; Peña-Gallardo et al., 2019). A recent study by Peña-Gallardo et al. (2019) focused on the responses of wheat and barley cropping systems to different drought indices over Spain, have shown the different efficacies of several drought indices, stressing the importance of the multiscalar character of droughts, in particular of the SPEI. Similarly, and in accordance to previous work by the authors (Ribeiro et al., 2019b), the present study shows the adequacy of SPEI for the assessment of the agricultural risks associated to droughts in the IP, and advances the added value of using the remote sensing of vegetation.

Overall, the results of the estimated copula functions have shown that Archimedean copulas are suitable to model the joint behaviour of yield anomalies and droughts , suggesting a dependence between extreme values of rainfed cereal yield anomalies and drought indicators, and the subsequent simulated distributions of crop yield anomalies are quite consistent with the observations. The results highlighted that the use of copulas for probabilistic assessment allow the estimation of the dependence in the tails of the distribution and were able to give the likelihood of crop-loss under drought conditions. This feature is of the most interest in risk analysis given that it models the joint probability of occurrence of crop-loss and drier events. Moreover, this study suggests the relevance of impact-centric approaches (also referred in literature as ‘bottom-up’ approaches (Zscheischler et al., 2018) to identify and characterize the hazards which lead to the larger impacts.

Moreover, it is important to stress that crop anomalies decline much more when drought conditions are below the mild or moderate threshold, suggesting a high agricultural drought risk level of wheat and barley in both clusters. While values of PNE the crop-loss threshold were low and similar for wheat in cluster 1 and barley in cluster 2, the values of CPNE the crop-loss threshold during drought years are considerably larger. The higher probability of crop-loss obtained when analysing only drought conditions agrees with Páscoa et al. (2017b), which have shown a very high agreement between low wheat yield anomalies and drought conditions in the IP, even on provinces where the linear correlation is no-significant.

Although there is a greater risk of crop-loss during drought conditions, some losses can still be expected during non-drought events, particularly in cluster 2 (14.1% and 7.77% in the case of wheat and barley, respectively). In the northern sector (cluster 1) the probabilities of crop-loss under non-drought conditions have the lower values, displaying 3.97% in the case of wheat and 3.65% in the case of barley. Some studies point to crop damages attributable to excessive wet soils (Rosenzweig et al., 2002; Zampieri et al., 2017) due to delayed planting or later harvest, nutrients runoff, development of pests and diseases, among others, highlighting the complexity of quantifying agricultural risk levels for management purposes, and the non-linear relation between crop yield and climate conditions. The lower values of CPNE under non-drought conditions in cluster 1 support the fact the slightly high values of PNE in cluster 1 are mainly dominated by drought conditions.

With the present study is not possible to establish sharp conclusions about the adequacy of the copula models to a specific type of drought indicator (remote sensing or hydro-meteorological), since only one type of drought indicator was considered for each cereal. In contrast, Bokusheva et al. (2016) have found that Gumbel copulas provided better fits representing the joint distribution of VCI and wheat, while Frank copulas described better the dependence between TCI and wheat yields, in Kazakhstan. Madadgar (2017) modelled the conditional probability density functions of crop yields under wet and dry conditions using SPI and SSI and found that a Clayton copula was the best function to model the dependence structures. Similarly, Leng and Hall (2019) have also used the same copula families and found that from 10 countries 5 of them featured Clayton copulas to fit the joint distribution between wheat production and SPI. However, the referred studies were somehow more restrictive as they do not take advantage

from using both remote sensing and hydro-meteorological drought indicators, and do not select the most important one *a priori*.

To further the research, the application of SPEI methodology to climate projections of precipitation and temperature holds an added-value to estimate drought risk levels for the next century. Likewise, the use of seasonal drought forecasts is also quite plausible in the approach presented in this study. Nevertheless, the presented results indicated the likelihood of crop-loss based on drought conditions observed much earlier than the harvest time, particularly in cluster 2 using SPEI (February and April with 5 and 1 month of time-scale). Hence, given the uncertainty associated to the seasonal forecasts for regional drought predictability in the IP, the use of past information for predictability studies is still successfully used (Pires and Ribeiro, 2016) and continues to be a source of information from an operational point of view. Other potential usefulness of this methodology for future research is the evaluation of its suitability at the province level and the assessment of whether other hazards (such as heat waves) are amplifying the impact of droughts on crop's harvest.

3.5 Conclusions

The agricultural drought risk levels estimated in the present work aimed to improve the effectiveness of the agricultural management of rainfed cereals in the major agricultural areas of the IP. The main findings of this study are summarized below:

- The dependence structure between crop yield anomalies and drought conditions is mainly asymmetrical, suggesting the existence of dependence among extreme values of yield anomalies and drought indicators.

- The differences between the unconditional and the conditional probability suggest that the risk of wheat-loss and barley-loss can be underestimated without conditioning the probabilities of non-exceedance crop thresholds to specific drought levels.

- The conditional probabilities of non-exceedance suggest that the risk of wheat-loss and barley-loss increases when drought events aggravate from normal or wet to moderate or severest conditions.

- The values of conditional probabilities of crop-loss under dry conditions suggest that the risk of drought-related barley-loss is more likely to occur in the northern sector, while the risk of drought-related wheat-loss is more likely in the southern sector, suggesting that sowing in cluster 1 (cluster 2) could be more focused on wheat (barley).

- The overall results show the importance of the concept of conditional probability to distinguish different meteorological settings associated to crop-losses and the applicability of the copula theory. The use of copula simulations for the analysis of the co-occurrence of dry and low-yield extreme events have shown the additional value of this methodology for the estimation of drought-related crop-failure.

- Nevertheless, minor wheat- and barley-losses can still be expected during normal or wet conditions, stressing the complexity of the interactions between the agricultural systems and the climate. Particularly, under the current climate change context, further high-impact-centric analysis are required, involving the cascading effects of different climate hazards.

Chapter 4

Copula-based agricultural drought risk of rainfed cropping systems

***Ribeiro, A.F.S.**, Russo, A., Gouveia, C.M., Páscoa, P. (2019). Copula-based agricultural drought risk of rainfed cropping systems, Agricultural Water Management, 223, 105689, https://doi.org/10.1016/j.agwat.2019.105689*

***Andreia F. S. Ribeiro contributions**: preparation of the data (crop yields, remote sensing and SPEI); statistical analysis and interpretation of the results (copula-based analysis and estimation of conditional probabilities); production of all the figures and coordination of the manuscript writing.*

Graphical abstract

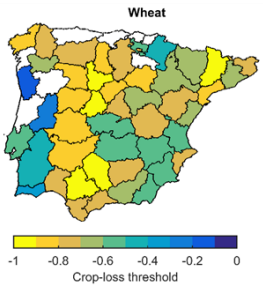
Data

Drought hazard

Remote sensing
VCI, TCI and VHI
Multiscalar drought index
SPEI

Drought impacts

Rainfed cereal yield
Wheat and barley t/ha
Crop loss events
Computed as -1 std of the crop yield on each province



Methods

1. Selected drought indicator for each province

Stepwise regression
Significance level of 0.05

2. Fit bivariate copula models

Copulas
Gaussian, t, Clayton, Frank and Gumbel
Copula selection
AIC

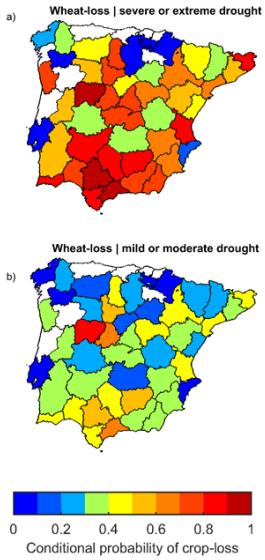
3. Map agricultural drought risk

Conditional probability of drought-related crop-loss at province level

Values ranging between **0** (**low** agricultural drought risk) and **1** (**high** agricultural drought risk)

VCI and TCI	SPEI	Drought class
>40	>-0.84	No drought
≤40	≤-0.84	Mild and moderate
≤20	≤-1.28	Severe and extreme

Results



Probability of **crop-loss increases** with drought severity in most of the provinces

Highlights

- Drought-related crop-loss in rainfed cropping systems was mapped for the Iberian Peninsula at the province level
- Most of the provinces exhibit an association between drought conditions and yield anomalies characterized by greater probabilities of joint low extremes, which is more pronounced in the case of barley
- In the case of the most selected type of copula (Clayton) drought conditions are mainly described by VCI and SPEI
- Regarding wheat, the second most selected type of copula (Gumbel) shows higher probabilities of joint high extremes, and drought conditions are mainly characterized by TCI in this case
- In most of the provinces, the probability of crop-loss increases with drought severity, with the probability of barley-loss being slightly higher than the probability of wheat-loss
- Some losses are still expected under non-drought conditions, stressing the complexity of the interactions between the agricultural systems and the climate

Abstract

In a future climate, warmer and drier conditions are expected, and the associated negative impacts in agricultural productions are a major issue. Assessing the risk of drought hazard on agricultural systems is, therefore, of main importance in decision-making, with the aim of mitigating drought-related crop-losses. In this study the agricultural drought risk is defined as the conditional probability of occurring crop-losses under drought conditions. We use the copula theory to estimate joint probability distributions describing the amount of dependence between drought conditions and crop yield anomalies of two major rainfed cereals in the Iberian Peninsula (wheat and barley), in the period 1986-2016. Further conditional probability distributions of the crop yield anomalies under different drought levels are obtained using the Standardized Precipitation Evapotranspiration Index (SPEI) and the satellite derived indices Vegetation Condition Index (VCI) and Temperature Condition Index (TCI). The results suggest that, in general, the joint behaviour of yield anomalies and drought conditions exhibits a dependence between the extreme values, whereas barley exhibits greater probabilities of joint extreme low values of yield and drought indicators. Moreover, while TCI is mainly used in copula models indicating greater probabilities of joint extreme high values of wheat and drought indicators (Gumbel models), VCI and SPEI are mainly associated to copula models indicating greater probabilities of joint extreme low values (Clayton models). The estimated conditional probabilities of occurrence of crop-loss are illustrated at the province level and suggest that agricultural drought risk increases with drought severity in most of the provinces.

Keywords: Drought; Rainfed cereals; Joint probability; Probability of crop-loss; Probabilistic decision-making

4. Copula-based agricultural drought risk of rainfed cropping systems

4.1 Introduction

Extreme weather events, such as droughts, are major sources of risk to agricultural systems mainly in rainfed conditions (Ben-Ari et al., 2016; Hernandez-Barrera et al., 2017; Páscoa et al., 2017b; Zampieri et al., 2017; Ribeiro et al., 2019b). The agricultural drought risk characterized by the probability of crop failure under the impact of dry conditions often entails substantial economic losses (Skakun et al., 2016; Xie et al., 2018). Hence, the combined assessment of the drought hazard and the associated impacts on crop production, based on a probabilistic approach, is suitable to capture the multivariate character of drought risk in agriculture (Madadgar et al., 2017).

To overcome drought identification and characterization problems, several approaches have been developed in the last decades. Among multivariate analysis approaches, copula functions are becoming quite popular (e.g. Mirabbasi et al. 2012; Lee et al. 2013; Li et al. 2015). Copulas are based on Sklar's theorem, linking one-dimensional marginal distributions to form a joint distribution. The joint distribution can be described by the respective marginal distributions and the copula which describes the dependence structure. Copulas have been applied to hydrological extremes (i.e., droughts and floods) focusing on the characteristics of the events, such as intensity, duration and frequency. Namely, Mirabbasi et al. (2012) estimated the bivariate joint behaviour of drought duration and severity, and a trivariate approach was adopted by Madadgar and Moradkhani (2013) for the joint behaviour of drought intensity, duration, and frequency under climate change scenarios.

Apart from being able to model two different characteristics of the same event (e.g. severity and duration of droughts), copula-based models can also be used to estimate the joint distributions of two or more different variables (e.g. Madadgar et al. 2017; Zscheischler and Seneviratne 2017). For instance, the recent and challenging understanding of the concurrent occurrence of different climate extremes, referred to as a compound event, have been addressed based on copula methods (Zscheischler and Seneviratne, 2017; Manning et al., 2018). Zscheischler and Seneviratne (2017) have addressed the structure of the dependence between

warm season temperature and precipitation using copulas, and they found that the co-occurrence of extremely hot and dry conditions is expected to increase under a strong greenhouse-gas forcing scenario.

The application of copula theory to agrometeorological studies is also relatively recent. The use of copula-based models to model the joint distribution of rainfed agricultural crops and drought conditions has been recently performed by e.g. Madadgar et al. 2017 for Australia and Bokusheva et al. 2016 in Kazakhstan. The previous mentioned authors have focused on establishing conditional probability distributions of yield in terms of copulas, characterizing droughts based on hydro-meteorological (Madadgar et al., 2017) and remote sensing data (Bokusheva et al., 2016), respectively. The appeal of using copulas in multivariate modelling is that the individual variables do not necessarily have the same probability distribution (Nelsen, 2006). In addition, the statistical inference of copulas is not restricted to methods implying the parametrical estimation of the margins (Maity, 2018).

In this work we use the unique abilities of copulas to link crop yield anomalies with drought information to achieve two main objectives: (1) estimate joint distributions (two-dimensional space) describing the amount of dependence between yield anomalies and drought conditions; and (2) use the estimated joint distributions to generate random crop yield anomalies preserving the structure of the dependence with droughts. The agricultural drought risk is evaluated in terms of the conditional probability of non-exceedance of the crop-loss thresholds, ranging between 0 (low agricultural drought risk) and 1 (high agricultural drought risk), to use a clearly and interpretable measure for decision support in case of agricultural drought. Here, drought condition is characterized using the remotely sensed Vegetation Condition Index (VCI) and Temperature Condition Index (TCI), and the hydro-meteorological multiscalar Standardized Precipitation Evapotranspiration Index (SPEI). In addition, drought impacts are assessed in terms of two major world crop productions (wheat and barley) particularly significant in the Mediterranean regions (Leff et al., 2004), where dryness is expected to increase in coming decades (Giorgi and Lionello, 2008).

4.2 Data and Methods

Data on crop yields were obtained from the Portuguese National Statistics Institute (INE) and the Spanish Agriculture, Food and Environment Ministry at the sub-national scale. We have

considered two winter cereals for analysis, wheat and barley, as they are major crop productions growing in rainfed conditions in the IP, and therefore susceptible to suffer from drought conditions (Austin et al., 1998; Vicente-Serrano et al., 2006; Páscoa et al., 2017b). Figure 4.1 illustrates the locations and names of the 54 considered provinces of Portugal (7 regions, referred as provinces hereafter, instead of regions, for sake of consistency with Spain nomenclature) and Spain (47 provinces). Three Spanish provinces (Vizcaya, Guipúzcoa and Asturias) were excluded due to the lack of production of one or both cereals on a consistent manner during the study period. Namely, the province of Vizcaya (Spain) has not produced barley during the analysed period, and it only produced wheat during 12 non-consecutive years with relatively small harvested area (between 1 and 5ha). In a similar way, we also excluded the province of Guipúzcoa (Spain) which has not produced wheat and barley during this period, and Asturias (Spain) which has not produced barley.

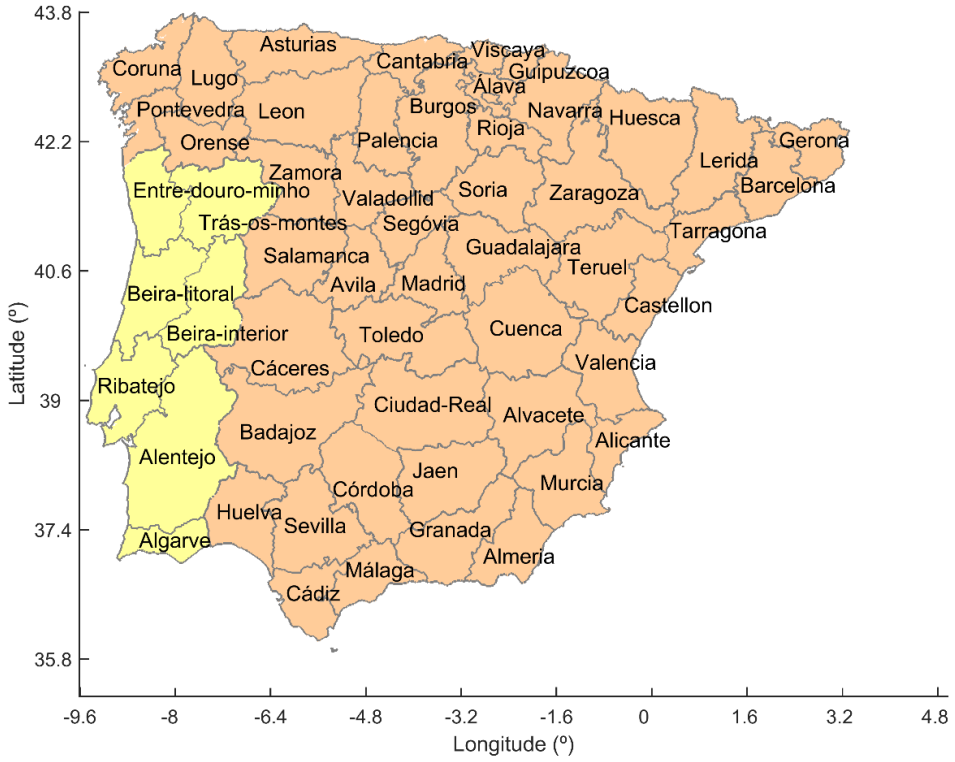


Figure 4.1 - Locations and names of the 54 considered provinces of Portugal (yellow) and Spain (orange provinces).

The annual values of wheat and barley yields (th^{-1}), were calculated as the ratio between crop production (t) and harvested area (ha) during the period of 1986–2016 for each province. The yield anomalies were computed by removing the linear trend, to exclude non-climatic factors (Gouveia and Trigo, 2008; Páscoa et al., 2017b). In this study we define crop-loss as events

when yield anomalies are below minus one standard deviation (-1std) of the time-series of annual crop yield anomalies. In this way, the target of our analysis is focused on estimating the expected chance that the negative yield anomalies will not exceed (i.e. is not higher than) -1std of each province and cereal mean yield, considering the average yield anomalies and the outliers, and not only the average (which would be the case of using zero as threshold and crop-loss events would correspond to yield anomalies only below zero and not below -1std). Figure 4.2 summarizes the yield standard deviation of the yield anomalies on each province and cereal. Notice that Portuguese provinces, particularly Entre-Douro-e-Minho, Beira-Interior and Algarve, show relatively low values of standard deviation, suggesting that crop-loss thresholds correspond to values slightly below normal, in comparison to Spanish provinces. Hence, in general, Portuguese annual yields are close to mean values and the difference between maximum and minimum yield values is small. On the other hand, Valladolid, Avila, Cordoba, Sevilla, Orense, Cuenca and Lerida, show the highest values of crop-loss threshold (Figure 4.2).

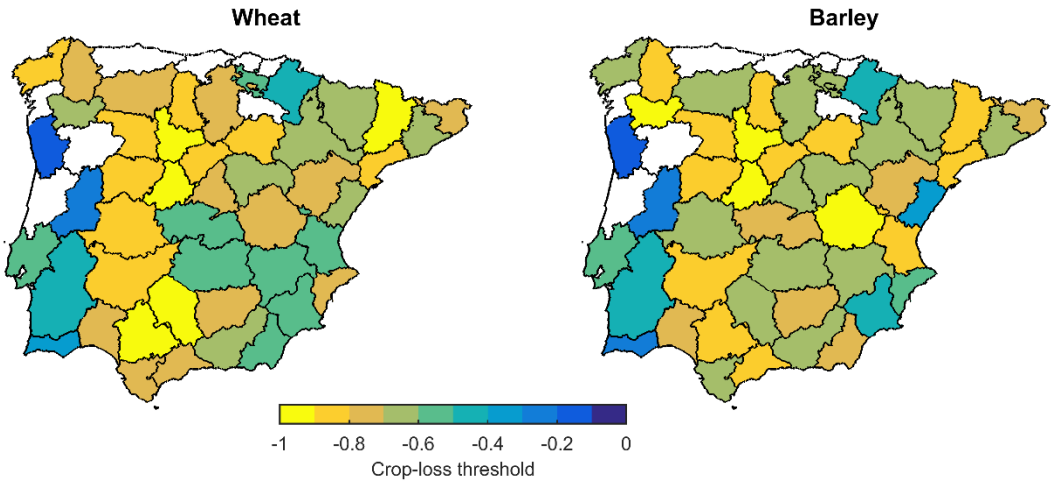


Figure 4.2 - Crop-loss threshold, computed as -1 std of the crop yield on each province. The provinces with no yield available data (Guipúzcoa, Vizcaya and Asturias) or with any statistically significant drought indicator (Trás-os-Montes, Beira-Litoral, Pontevedra, Asturias and Rioja) are coloured in white.

In this study, drought hazard was evaluated using remotely sensed and hydro-meteorological multiscalar drought indices. In terms of remote sensing, we employed the satellite-based Vegetation Condition Index (VCI) (Kogan et al. 2015) and the Temperature Condition Index (TCI) (Kogan 1995). VCI and TCI monitor vegetation condition based respectively on the Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT), being indicators of moisture and thermal stress on vegetation, respectively. Weekly global maps of VCI and TCI were retrieved at 4 km spatial resolution from NOAA's ftp server

(ftp://ftp.star.nesdis.noaa.gov/pub/corp/scsb/wguo/data/Blended_VH_4km/geo_TIFF/), during 1985–2016, and spatial averages of these indices were computed for each province.

In terms of hydro-meteorological drought indices, we used the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) at different time-scales (1 to 12 months) during 1986-2016. The SPEI computation considers the role played by the evaporative demand on the occurrence of dry events, consisting in a monthly climatic water balance (precipitation minus reference evapotranspiration) (Vicente-Serrano et al., 2010). In this study, the monthly drought index SPEI gridded values, with spatial resolution of 0.5°, were computed based on precipitation and reference evapotranspiration values from the Climate Research Unit (CRU TS4.01), using a log-logistic distribution for statistical fitting (Vicente-Serrano et al. 2010), and a gaussian kernel function to accumulate the previous months. Further spatial averages were computed for each province.

Similar to earlier work performed by the authors (Ribeiro et al., 2019b), the indices VCI and TCI were examined between the weeks 35 and 25, comprising the major crop life cycle moments, and the indices SPEI at 12 time-scales were examined from January to June (due to the time scale memory). However, in the present work we extended the data to the period 1986-2016 and to the province level, while Ribeiro et al. (2019b) considered the shorter period 1986-2012 and only 2 clusters of 5 and 4 provinces of the IP. This high spatial scale is likely to better reflect the high spatial variability of climate conditions in the IP (Esteban-Parra et al., 1998; Trıgo and Dacamara, 2000; Beguería and Vicente-serrano, 2006; Trigo et al., 2013). The geographic distribution of mean rainfall amounts over the IP is illustrated in Figure 4.3 for the vegetative cycle of the winter crops (from sowing approximately in September of the year n – 1 to the harvest approximately in June of the year n), based on the precipitation dataset considered to compute the SPEI. The precipitation regime exhibits strong spatial gradients, with higher values in the northwestern sector and lower values in the southeastern sector as reported by several authors (Belo-Pereira et al., 2011; Andrade and Belo-Pereira, 2015). The northwestern-southeastern contrasts in the precipitation regime are responsible for the spatial distribution of the vegetation behaviour in the IP, and for these reason Gouveia et al. (2017) have also found a similar gradient of vegetation clusters.

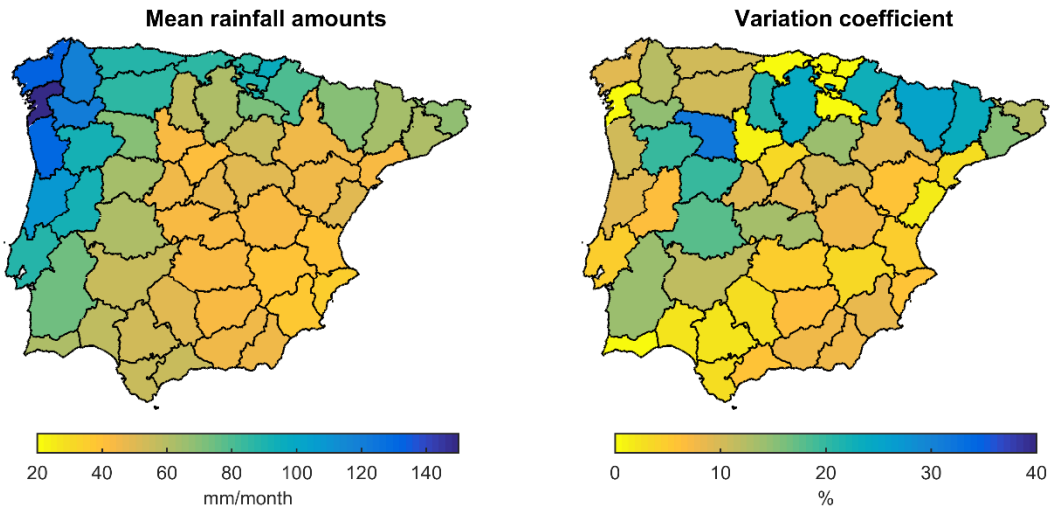


Figure 4.3 - Spatial distribution of mean rainfall amounts (mm/month) over the Iberian Peninsula (IP) at the province level from September to June for the 1986–2016 period (left) and percentage of variation coefficient of mean rainfall amounts (right).

The dominant drought indicator in explaining each cereal time-series variability was selected for each province (Table 4.1). The baseline pool of drought indicators includes 43-week intervals of VCI, 43-week intervals of TCI, and 6-month by 12-time-scales intervals of SPEI, totalling 158 possible drought indicators for each cereal and province. A stepwise linear regression was applied to perform a prior selection of the best subset of drought indicators for each cereal, followed by the selection of the one with largest absolute value of the standardized regression coefficients. Table 4.1 indicates the selected drought indicator for each provinces and respective selected week (in the case of VCI and TCI) or month and time-scale (in the case of SPEI). According to Table 4.1, all the selected drought indicators are statistically significant at the 5% significance level (the null hypothesis tests if the corresponding regression coefficient is equal to zero, under the influence of the remaining drought indicators selected by stepwise regression). In order to assess how each selected drought indicator influences crop's variability we use partial correlation. The squared partial correlation coefficient indicates how much of the variance which is not estimated by the remain drought indicators chosen by stepwise regression is estimated by the selected drought indicator. In the results section we will discuss in more detail the selected drought indicators (Figure 4.4).

Table 4.1 - Selected drought indicator and for each province (significance level of 0.05) and respective p-value. The provinces with no yield available data (Guipúzcoa, Vizcaya and Asturias) or with any statistically significant predictor (Trás-os-Montes, Beira-Litoral, Pontevedra, and Rioja) are denoted by '- '. In the case of SPEI, the numbers correspond to the selected month and time scale ('SPEI-month-timescale'), and in the case of the remote sensing indices the numbers correspond to the selected week ('VCI-week').

	Provinces	Wheat		Barley	
		Drought indicator	p-value	Drought indicator	p-value
Portu	Entre Douro e Minho	SPEI-1-2	0.001	VCI-35	0.029
	Trás-os-Montes	-	-	-	-

	Beira-litoral	-	-	-	-
	Beira-interior	SPEI-3-4	0.0004	SPEI-3-10	0
	Ribatejo-oeste	TCI-48	0.0155	VCI-17	0.0033
	Alentejo	VCI-14	0	VCI-14	0
	Algarve	VCI-18	0.0022	VCI-16	0
Spain	Coruna	SPEI-2-12	0.039	TCI-50	0.0178
	Lugo	VCI-1	0.0318	SPEI-5-2	0.0332
	Orense	SPEI-1-1	0.0395	VCI-12	0.0413
	Pontevedra	-	-	-	-
	Asturias	-	-	-	-
	Cantabria	-	-	-	-
	Alava*	TCI-24	0.0073	VCI-20	0.0099
	Guipúzcoa	-	-	-	-
	Vizcaya	-	-	-	-
	Navarra	TCI-9	0	TCI-10	0.0005
	Rioja	-	-	-	-
	Huesca	TCI-22	0.0011	SPEI-3-8	0
	Teruel	SPEI-6-11	0.0001	SPEI-6-10	0
	Zaragoza	SPEI-4-10	0.0011	SPEI-5-4	0
	Barcelona	SPEI-3-10	0	SPEI-3-10	0.0002
	Gerona	SPEI-1-6	0.0001	TCI-25	0
	Lerida	TCI-25	0.0021	VCI-14	0.0041
	Tarragona	SPEI-4-10	0.0001	SPEI-6-11	0.0001
	Avila	VCI-25	0	SPEI-3-5	0
	Burgos	VCI-21	0	TCI-23	0
	Leon	TCI-19	0.0001	SPEI-2-11	0.0001
	Palencia	VCI-25	0.0037	TCI-23	0
	Salamanca	VCI-24	0.0005	VCI-24	0.0011
	Segovia	SPEI-2-1	0.003	SPEI-3-5	0.0002
	Soria	TCI-22	0	SPEI-6-12	0
	Valladolid	VCI-24	0.0001	VCI-25	0.001
	Zamora	TCI-21	0	VCI-20	0
	Madrid	SPEI-3-11	0.0014	VCI-18	0.0019
	Albacete	TCI-48	0.0003	SPEI-3-5	0
	Ciudad-real	TCI-46	0.0003	SPEI-3-5	0
	Cuenca	SPEI-6-11	0	SPEI-6-11	0
	Guadalajara	SPEI-3-11	0.0004	SPEI-4-5	0
	Toledo	SPEI-6-11	0	SPEI-5-5	0
	Alicante*	TCI-36	0.0002	SPEI-6-9	0
	Castellon	SPEI-3-5	0.0007	VCI-23	0
	Valencia	SPEI-6-11	0	SPEI-6-5	0
	Murcia	VCI-13	0.017	SPEI-5-5	0
	Badajoz	VCI-10	0.0094	VCI-16	0
	Caceres	TCI-19	0.0242	VCI-19	0.0118
	Almeria	SPEI-6-12	0.0136	SPEI-6-11	0.0005
	Cadiz	VCI-19	0.0007	VCI-19	0.0019
	Cordoba	SPEI-5-1	0.0003	VCI-14	0.0002
	Granada	SPEI-5-10	0	SPEI-6-11	0
	Huelva	VCI-13	0.0002	VCI-13	0
	Jaen	SPEI-5-4	0.0037	SPEI-4-4	0.0052
	Malaga	VCI-10	0.0002	VCI-9	0.0004
	Sevilla	VCI-15	0.0012	SPEI-6-10	0

After the selection of one drought indicator for each cereal and province, bivariate models based on copula functions are estimated. Copulas were firstly mentioned in the literature more than half a century ago by Sklar (1959). However, their application to the environmental and Earth sciences is more recent (Maity, 2018). Mathematically, given two correlated variables X and Y

with marginal distributions $F_X(x)$ and $F_Y(y)$, uniform on the interval [0,1], a copula function *C* links these distributions to their joint probability distribution $F_{XY}(x, y)$ as follows:

$$F_{XY}(x, y) = C\big(F_X(x)\,, F_Y(y)\big) \tag{4.1}$$

Here, X denote the yield anomalies, Y a drought indicator (SPEI, VCI or TCI), and (x, y) are pairs of observations. A more detailed description of the use of copulas is provided by Nelsen (2013).

A total of five commonly used copulas were considered to estimate the bivariate joint dependence structures: Gaussian, student's t-copula, Clayton, Gumbel and Frank (Table 4.2). These well-documented copula functions belong to two distinct classes of copulas: Elliptical (Gaussian and student's t-copula) and Archimedean (Clayton, Gumbel and Frank). Among all types of copulas described in the literature, most of the recent studies using copulas for agrometeorological purposes have employed the classes referred above given their properties (Bokusheva et al., 2016; Madadgar et al., 2017; Zscheischler et al., 2017). The Archimedean copulas have an explicit formula with only one parameter (Table 4.2) and are quite popular given their ability in capturing a wider variety of joint dependence structures and, in some cases, joint tail dependence (i.e. joint extreme events). The Clayton and Gumbel copulas describe an asymmetrical tail behaviour exhibiting greater dependence in the lower and upper tail, respectively, while the students t-copula is only able to model joint symmetric tail dependence. However, the Frank copula, in a similar way to the Gaussian copula, is only capable to capture joint symmetric dependence without tail dependence. Despite the popularity of the Elliptical copulas because they derive from well-known distributions associated to the widely used Pearson's correlation, the elliptical dependence is only able to capture radial symmetry and the mathematical expressions do not have a closed form (Table 4.2).

Table 4.2 - Equations of the copula functions, where u and v are univariate variables, Φ^{-1} is the inverse of standard CDF, ${t_{df}}^{-1}$ is the inverse t–student CDF, df is the degree of freedom, ρ and θ are dependence parameters.

Family	**Joint Cumulative Distribution Function** $C\big(F_X(x)\,, F_Y(y)\big) = C(u, v)$	**Parameter range**
Gaussian	$\int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} e\left(-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right) dudv$	$-1 \leq \rho \leq 1$

t	$\int_{-\infty}^{t_{df}^{-1}(u)} \int_{-\infty}^{t_{df}^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} e\left(1+\frac{u^2+v^2-2\rho uv}{df(1-\rho^2)}\right)^{-\frac{df+2}{2}} dudv$	$-1 \leq \rho \leq 1$ $df \geq 1$
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}}$	$0 \leq \theta$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$\theta \neq 0$
Gumbel	$e^{[(-lnu)^{\theta}+(-lnv)^{\theta}]^{\frac{1}{\theta}}}$	$1 \leq \theta$

The mentioned copulas were tested and selected based on the smallest value of Akaike's Information Criteria (AIC) (Li et al. 2015), for each cereal and province. In order to transform the individual variables to uniformly distributed values between 0 and 1, as required by copula definition, we used the kernel density estimator of the Cumulative Distribution Function (CDF). This procedure is named Canonical Maximum Likelihood (CML) method (Genest et al., 1995), which allows for semi-parametric estimation of the copula models. Using this procedure, the estimation of the margins is performed non-parametrically and the later fitting of the copula functions is performed by means of maximum likelihood to determine the copula parameters (Genest et al., 1995). The main aim of adopting a semi-parametric methodology is to avoid heavy assumptions about the marginal distributions when the sample size is rather small, which can compromise the robustness of the models. The data sample dimension n=31 used in this study results from the annual nature of the harvest values and from the use of remote sensing information that began in the eighties of the last century. Nevertheless, the use of a semi-parametric methodology is a good alternative without sacrificing the use of an important source of information complementary to climate variables, such as the remote sensing of vegetation.

The copula parameters were estimated, and afterwards, uniformly distributed data was simulated, originating $F_{X_{sim}}(x)$ and $F_{Y_{sim}}(y)$, which stand for simulated distributions of yield anomalies and drought indicator, respectively. The data generation simulated 1000 pairs of data in the range [0, 1], which were transformed back to the original scale using the kernel estimations of the inverse CDF. This way, we obtained simulations of yield anomalies (X_{sim}) and respective drought indicator (Y_{sim}), and selected the data points of X_{sim} which correspond to particular drought conditions (y^*) of Y_{sim} (severe or extreme droughts, mild or moderate drought and no-drought – see Table 4.3) such that

$$X_{sim^*}(x, y) = X_{sim}(x|y < y^*) \quad (4.2)$$

As follows, the conditional probability distributions of yield anomalies for different drought levels $F_{X_{sim^*}}$ preserve the structure of the dependence between the original marginal distributions of yield anomalies $F_X(x)$ and the drought indicator $F_Y(y)$ by using their joint distribution. For each drought condition, the risk of crop-loss was evaluated in terms of the Conditional Probability of Non-Exceedance (CPNE) of the crop-loss thresholds of each cereal and province (Figure 4.2), given by

$$F_{X_{sim^*}} = (-X_{std}) = Pr(X_{sim^*} \leq -X_{std}) \quad (4.3)$$

where $-X_{std}$ is the crop-loss threshold (-1std) for each province and cereal. In this way, in this study we define the agricultural drought risk as the conditional probability of occurring crop-losses given that certain drought conditions are met.

Table 4.3 - Drought class severity adapted from Kogan (2002) and Agnew (2000) for VCI/TCI and SPEI, respectively.

VCI and TCI	SPEI	Drought class
>40	>-0.84	No drought
≤40	≤-0.84	Mild and moderate drought
≤20	≤-1.28	Severe and extreme drought

4.3 Results

Drought conditions were characterized based on one drought indicator for each cereal (wheat and barley) at the province level (Figure 4.4 and Table 4.1), according to the absolute value of the standardized regression coefficient. Figure 4.4 shows the selected drought indicator (VCI, TCI or SPEI) and the variance of the crop yield explained by this indicator, (although not shown, the corresponding partial correlations are all statistically significant). According to Figure 4.4a, the SPEI is selected to characterize drought conditions in almost half of the provinces in the IP (20 provinces in the case of wheat and 23 in the case of barley), followed by the VCI (14 provinces in the case of wheat and 18 in the case of barley), and TCI (12 provinces in the case of wheat and 5 in the case of barley). While SPEI is mostly selected in the eastern provinces of the peninsula, VCI is mostly selected in western regions, in the case of both rainfed cereals. The TCI is the drought indicator less selected, presenting its influence on the northern provinces of Spain in both cereals, and also in central-south provinces of the IP in the case of wheat.

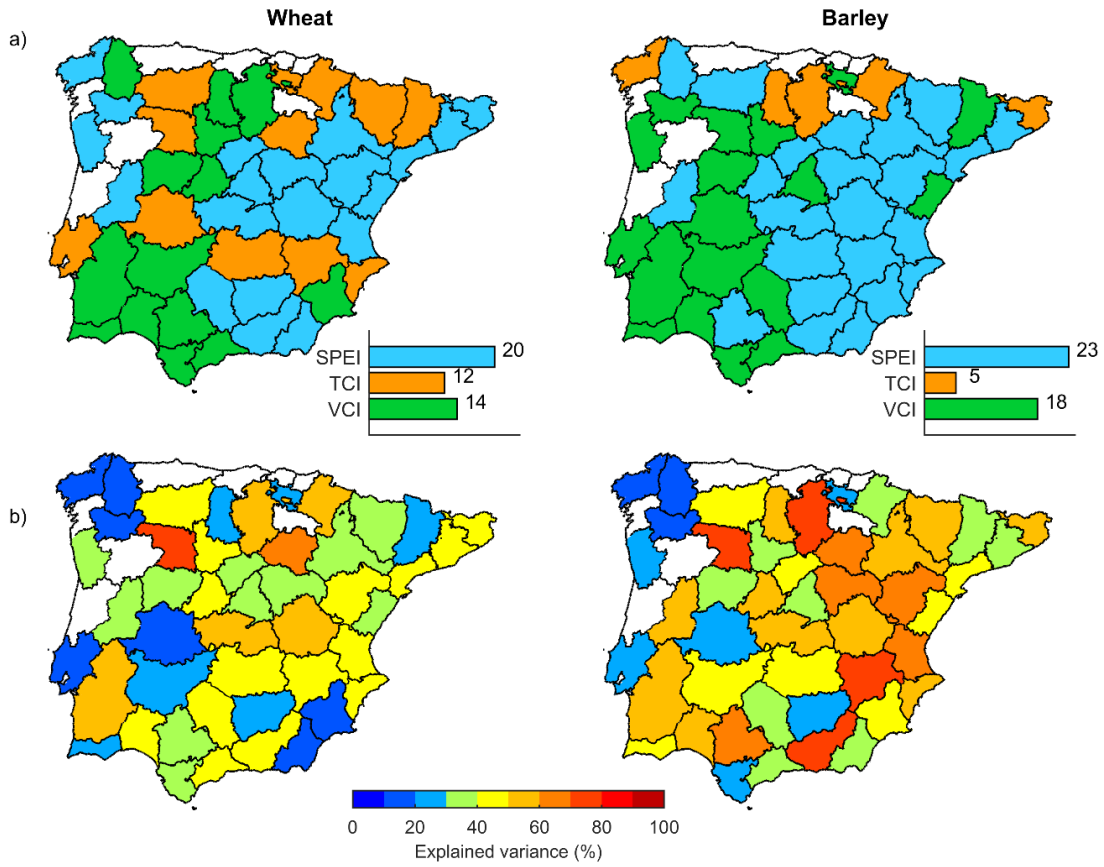


Figure 4.4 – a) Type of predictor (VCI – green, TCI – orange, or SPEI – blue) selected for each cereal and province and the respective number of provinces using each drought indicator (small bar graphs); b) variance of each cereal yield variability explained (%) by the selected drought indicator. The provinces with no yield available data (Guipúzcoa, Vizcaya and Asturias) or with any statistically significant predictor (Trás-os-Montes, Beira-Litoral, Pontevedra, and Rioja) are coloured in white.

In general, the provincial barley yield shows a larger number of provinces with greater explained variance, in comparison to wheat (Figure 4.4b). Moreover, both cereals show most of lower values of explained variance in northwestern provinces, while higher values of explained variance are mainly present in eastern provinces. According to Figure 4.4b more than 70% of the yield variability in Zamora is explained in both cereals, as obtained by the standardized regression coefficients, and the correspondent drought conditions are characterized by the TCI (week 21, middle of May) in the case of wheat, and by the VCI (week 20, beginning of May) in the case of barley (see also Table 4.1). Besides Zamora, the provinces Burgos, Albacete and Granada also show a percentage of explained variance higher than 70% in the case of barley (Figure 4.3). While drought conditions influencing barley yield in Burgos (northern province) are characterized by TCI (week 23, end of May), the drought conditions in Albacete and Granada (southeastern provinces) are characterized by the SPEI (March with 5-month time-scale and June with 11-month time-scale, respectively – see Table 4.1). Ribeiro et al. (2019b) also showed that there are stronger relationships between remote sensing indices

and cereal yield in the western regions of the IP, while the yield variability in the eastern region is better predicted based on SPEI.

In Portugal, Alentejo is the province with the highest explained variance of cereal yield variability, and the drought conditions are characterized by the VCI in the week 14 (end of March) in both cereals (Figure 4.4b). As a matter of fact, Alentejo is the province producing most of the rainfed cereals in Portugal (Gouveia and Trigo 2008) and it is particularly affected by drought episodes (Moreira et al., 2013).

The fitting of the copula functions from Table 4.2 (Gaussian, t-copula, Clayton, Frank and Gumbel) was carried out for each pair of cereal yield and drought indicator. Figure 4.5a maps the type of selected copula according to the lowest value of AIC, Table S4.1 (in Supplementary Material) indicates the respective dependence parameter estimates, and Figure 4.5b illustrates the number of provinces adopting each type of copula, and for each type of copula the number of provinces using each type of drought indicator.

In both cereals, the Clayton copula is the most selected copula type (Figure 4.5) (14 provinces in case of wheat and 21 in the case of barley), which indicates greater probabilities of joint extreme low values (i.e., lower values of yield anomalies given lower values of drought indicators). The opposite situation, i.e. greater number of provinces showing larger probabilities of joint extreme high values of yield anomalies and drought hazard (i.e., higher values of yield anomalies and higher values of the drought indicator) is also verified, corresponding to Gumbel models. However, while there are more Clayton copulas in the case of barley in comparison to wheat, there are more Gumbel models in the case of wheat rather than in barley (12 provinces Gumbel copula in the case of wheat and 8 in the case of barley). Moreover, in the case of wheat all the Clayton models have drought conditions characterized by VCI and SPEI, and half of the wheat Gumbel models uses TCI. In a similar way, in the case of barley only one of the 21 Clayton models uses TCI (hence 20 of the 21 Clayton models uses VCI or SPEI), but most of the Gumbel models uses SPEI, the majority of Clayton models uses VCI, and the number of provinces using TCI is much lower (see Figure 4.4a). As expected, the t-copula is the less selected type (3 provinces in the case of wheat and 1 province in the case of barley) given that AIC penalizes the number of estimated parameters and t-copulas require the estimation of two parameters (Table 4.2). Regarding the copula functions without asymptotical tail dependence (Gaussian and Frank) there are more Frank copula models in the case of wheat (10 provinces

in the case of wheat and 7 in the case of barley), and more Gaussian copula models in the case of barley (7 provinces in the case of wheat and 9 in the case of barley).

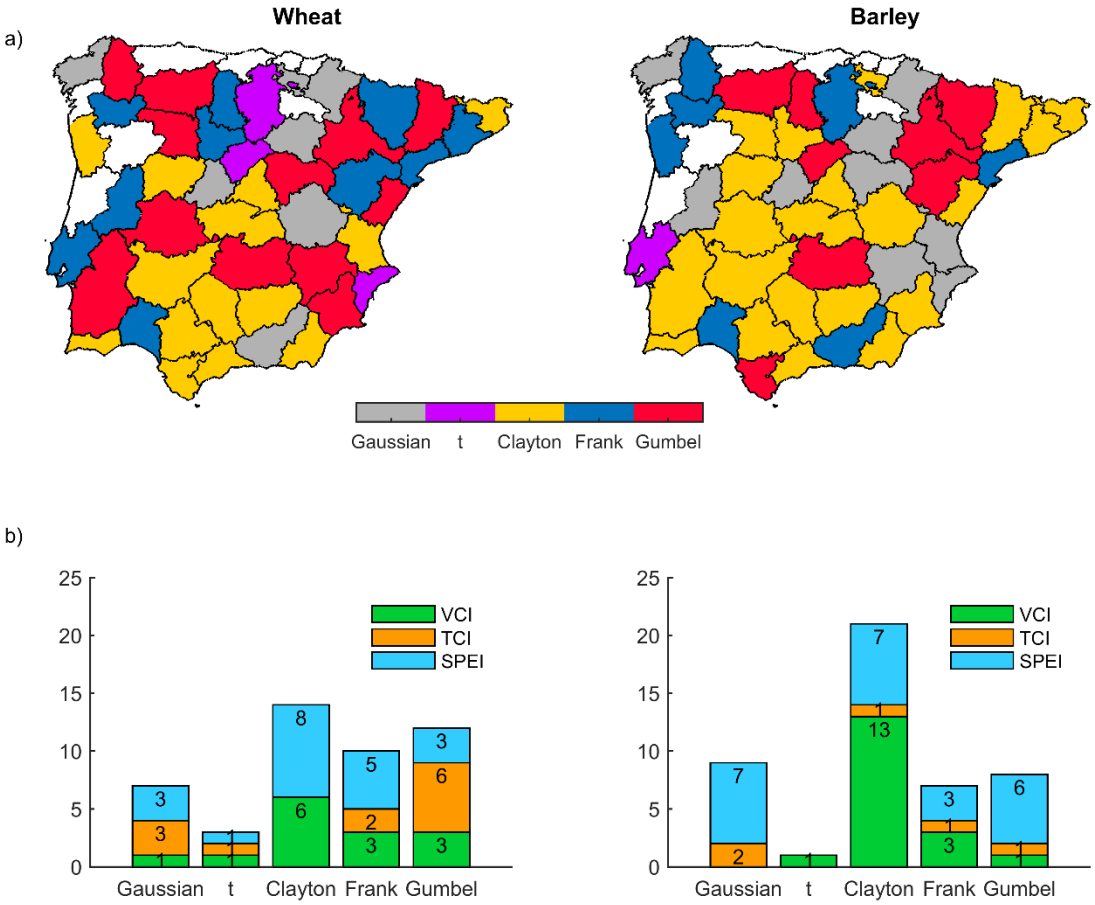


Figure 4.5 - Selected copulas according to the values of AIC. The provinces with no yield available data (Guipúzcoa, Vizcaya and Asturias) or with any statistically significant predictor (Trás-os-Montes, Beira-Litoral, Pontevedra, and Rioja) are coloured in white.

After estimating the joint probability distributions using copula functions, the subsequent generated simulations allowed to assess the conditional probability distributions of crop yield anomalies under different drought levels (no drought, mild or moderate drought and severe or extreme drought, see Table 4.3). In Figure 4.6 the conditional probability of crop-loss given different drought severity classes is shown based on the crop-loss thresholds mapped in Figure 4.2. Complementarily, Figure 4.7 indicates the number of provinces showing each decile of conditional probability of crop-loss, and the respective type of copula.

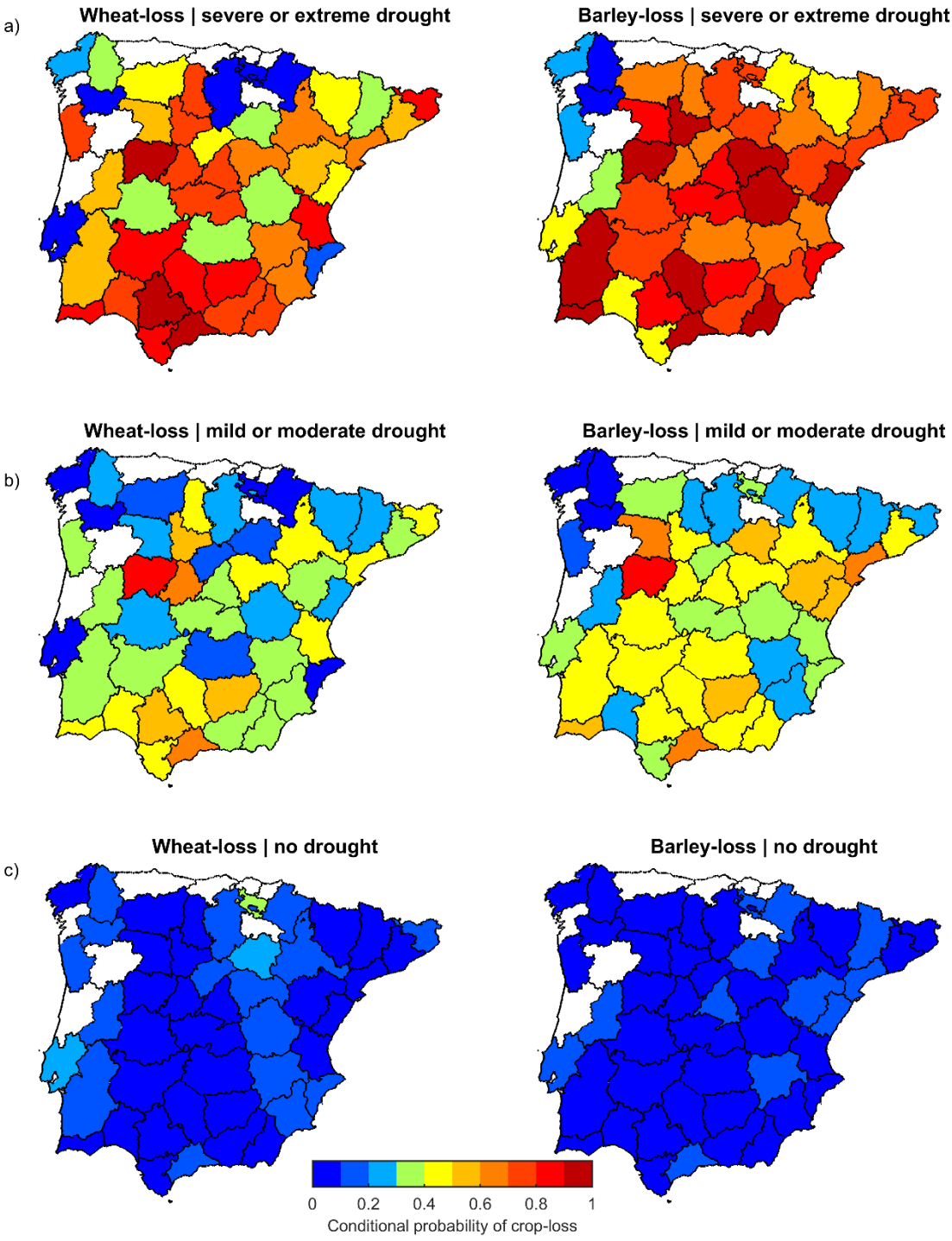


Figure 4.6 - Agricultural drought risk during a) severe or extreme droughts; b) mild or moderate drought; c) no-drought.

In general, the agricultural drought risk increases with drought severity in most of the provinces, with particular emphasis on the provinces of Salamanca and Malaga in the case of the two cereals. In comparison to the other provinces, Salamanca exhibits the greatest drought risk of wheat and barley during mild or moderate droughts (Figure 4.6b) and a slight increase during

severe or extreme droughts (Figure 4.6a). In general, in comparison to wheat during mild or moderate drought (Figure 4.6b), barley shows a larger number of provinces with more than 40% chance of non-exceedance of the respective crop-loss thresholds. During severe or extreme droughts (Figure 4.6a) the agricultural drought risk increases in most of the provinces, in comparison to mild or moderate droughts (Figure 4.6b), and barley endures a larger number of provinces with more than 70% chance of non-exceedance of the respective crop-loss thresholds, in comparison to wheat. During non-drought conditions (Figure 4.6c) the agricultural drought risk is considerably lower in most of the provinces, ranging from 0% to 30% chance of non-exceedance of the wheat-loss thresholds, and ranging from 0% to 20% chance of non-exceedance of the barley-loss thresholds.

Despite the general increase of the agricultural drought risk with drought severity, minor exceptions are detailed next. The provinces Ribatejo-oeste (Frank copula), Ourense (Frank copula), Alava (Gaussian copula), and Navarra (Gaussian copula) show a decrease of wheat drought risk with drought severity, decreasing progressively from no-drought to mild or moderate drought and to severe or extreme drought (Figure 4.5 and Figure 4.6). The provinces Segovia (t-copula), Soria (Gaussian copula) and Alicante (t-copula) also decrease wheat drought risk from the situation of non-drought to mild or moderate drought, while Burgos (t-copula) decreases wheat drought risk from the situation of mild or moderate drought to severe or extreme drought (Figure 4.5 and Figure 4.6). Regarding barley, only Lugo (Frank copula) and Ourense (Frank copula) provinces decrease agricultural drought risk from non-drought conditions to mild or moderate drought. All the provinces modelled with Gumbel or Clayton copulas show an increase of the agricultural drought risk with drought severity, while these exceptions are modelled by Gaussian, Frank or t-copulas (Figure 4.7).

As expected, Figure 4.7 indicates that the number of provinces in lower deciles of agricultural drought risk is higher during non-drought conditions, followed by a higher number of provinces in intermediate deciles during mild or moderate droughts, and the higher number of provinces in higher deciles is observed during severe or extreme droughts. In addition, the majority of the provinces with more than 70% chance of non-exceedance of the respective crop-loss thresholds, during severe or extreme droughts, is modelled by Clayton copulas in both cereals (Figure 4.7a yellow bars). In a similar way, the Clayton provinces are also distributed in the greater deciles of agricultural drought risk during mild or moderate droughts of both cereals (Figure 4.7b yellow bars). During severe or extreme droughts and mild or moderate droughts, the Gumbel

provinces are mainly associated to intermediate deciles of agricultural drought risk (Figure 4.7a-b red bars). The lower deciles of severe or extreme droughts and mild or moderate droughts are mainly associated to Gaussian, Frank or t-copulas (Figure 4.7 grey, purple and blue bars).

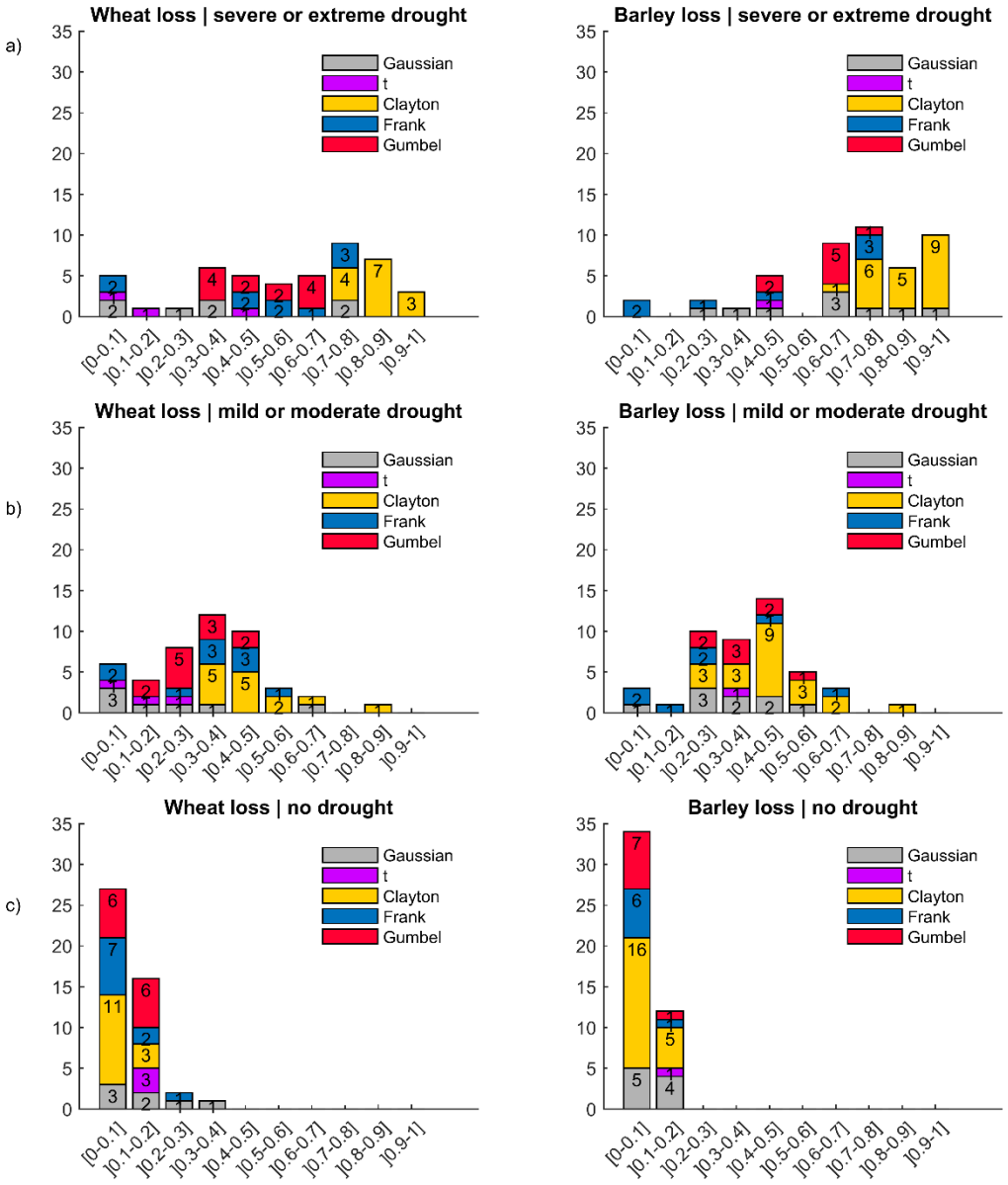


Figure 4.7 - Histogram indicating the number of provinces for each agricultural drought risk interval and respective type of copula.

4.4 Discussion and conclusions

This work aimed to apply the copula approach for estimating the amount of dependence between crop yield anomalies (wheat and barley) and drought conditions in terms of the joint cumulative distribution function. Based on the estimated dependence structures, the agricultural drought risk in the IP was defined as the conditional probability of non-exceedance of crop-loss

thresholds under different drought levels: no drought, mild or moderate drought and severe or extreme drought. The drought hazard was characterized in terms of the SPEI, VCI and TCI, combining the advantages of both hydrometeorological and remote sensing data for the assessment of drought conditions.

The results were illustrated at the province level of the IP and for each cereal, and intended to address the following key questions: 1) which are the critical drought indicators influencing crop yield variability? 2) which are the most adequate copula functions to represent the dependence between crop yield anomalies and drought indicators? 3) is there some type of copula most closely associated to the type of drought indicator? 4) what is the conditional probability of joint extreme low values, i.e., the likelihood of crop-loss under different levels of drought severity? 5) is there some type of copula most closely associated to higher values of agricultural drought risk? 6) which are the provinces and cereals with higher agricultural drought risk, according to the applied methodology? The following paragraphs address the fundamental findings of the present study and answer the previously established questions.

In general, the SPEI was found to be the drought indicator most adequate to the highest number of provinces in both cereals (Figure 4.4a). However, in comparison to wheat, barley presents a larger number of provinces characterized by SPEI and with larger explained variance (Figure 4.4). On the other hand, the provinces with higher explained variance do not necessarily coincide with the selection of SPEI. After the SPEI, the VCI is the second most selected type of drought indicator, particularly regarding barley yield anomalies, followed by TCI, (Figure 4.4a). As a matter of fact, in the case of barley we can observe a west-east gradient of suitability of VCI in the western provinces towards SPEI in the eastern provinces (Figure 4.4a). This west-east gradient is also slightly present in the case of wheat, thus concluding that the rainfed cereal yield variability in western part of the IP is largely explained by VCI, while the rainfed cereals variability in eastern part of the IP is mostly explained by SPEI. In a similar way, the spatial patterns of humid/semiarid climate regimes, precipitation patterns and type of vegetation exhibit a northwestern-southeastern gradient in the IP (Soriano and Pablo, 2001; Belo-Pereira et al., 2011; Gouveia et al., 2011). In other words, the results suggest that the greenness of vegetation is a better proxy of yield anomalies in western rainfed cereals (humid areas), whereas water balance is a better proxy of yield anomalies in the eastern rainfed cereals (semiarid areas). In previous works using NDVI instead of VCI (which is NDVI-derived) the authors have also found a strong relationship between the wheat yield and vegetation activity in a western Iberian

region (Alentejo) (Gouveia and Trigo, 2008). The influence of TCI in the northern wheat and barley yields (higher latitudes and elevations) and in central parts of the IP in the case of wheat, is also observed, suggesting the importance of solar radiation for vegetation growth in these regions (Bento et al., 2018).

The most selected type of copula in both cereals was found to be Clayton functions (Figure 4.5), suggesting stronger dependence between the extreme left values of drought conditions and yield anomalies, in most of the provinces. In addition, most of the Clayton models in the present work have drought conditions characterized by VCI and SPEI, indicating that the drought conditions characterized by these indicators greatly influence the occurrence of crop-loss. On the other hand, the Gumbel models (greater probability of joint extreme high values) in the case of wheat have most of drought conditions characterized by TCI, suggesting the influence of the increased solar radiation in the positive wheat anomalies in these provinces. Hence, the results suggest that high values of TCI promote positive yield anomalies of wheat (over 6 of the Gumbel provinces), while low values of VCI and SPEI promote negative yield anomalies of wheat (over 14 Clayton provinces) and barley (over 20 Clayton provinces).

In general, during severe or extreme drought, and mild or moderate drought, most of the provinces with greater deciles of probability of crop-loss are associated to Clayton copulas, while the lower deciles are more linked to copula functions without asymptotical tail dependence (Gaussian and Frank) or without asymmetrical tail dependence (t-copulas). Moreover, the provinces which do not increase agricultural drought risk with drought severity (Ribatejo-oeste (wheat), Ourense (wheat and barley), Alava (wheat), Navarra (wheat), Segovia (wheat), Burgos (wheat) and Lugo (barley)) are all modelled by copulas without tail dependence or t-copulas, thus suggesting that other factors rather than drought, could be the reason of crop failure in these regions. In a similar way, (Páscoa et al., 2017b) have also found negative correlations between wheat yield and SPEI in provinces where there is an agreement between dry events and low yield anomalies. In the present work, most of the provinces which do not increase probability of crop-loss with drought severity (with the exception of Burgos and Lugo) have drought conditions characterized by TCI or SPEI, suggesting that increased solar radiation and temperature could be favourable for cereal yield in these cases.

As expected, most of the higher probabilities of agricultural drought risk (during mild or moderate droughts and severe or extreme droughts) are observed in provinces modelled by Clayton copulas (Figure 4.7), which are closely associated to VCI and SPEI, suggesting that

these indices are potential good indicators of drought-related crop-failure. Moreover, to a greater severity of drought corresponds a higher agricultural drought risk in both cereals in most of the provinces. However, while barley exhibits 10 provinces with conditional probabilities of crop-loss under severe or extreme drought higher than 90%, wheat only exhibits 3. During mild or moderate droughts, barley also exceeds wheat in the number of provinces with 40% to 50% chance of crop-loss. Hence, the results support that barley exhibits greater agricultural drought risk, rather than wheat, in agreement with the fact that barley is the cereal with more provinces with joint behaviour of yield anomalies and droughts described by Clayton copulas.

Similarly to the patterns found in the spatial distribution of precipitation (Figure 4.3) and type of drought indicator (Figure 4.4), one would expect a similar behaviour between neighbouring provinces in terms of selected copula models and subsequent conditional probabilities of drought-related crop-loss. As a matter of fact, in some cases even diametrically opposed dependence structures (Clayton and Gumbel copulas) are selected by neighbouring provinces. This shortcoming may be associated to the distinct moments of the vegetative cycle selected to characterize droughts as shown in Table 4.1, and to the province- and crop-specific threshold adopted in the present study to define crop-loss events (Figure 4.2). Moreover, despite the spatial pattern in the mean rainfall in Figure 4.3a, the variation coefficient in Figure 4.3b also do not present an evident spatial pattern with neighbouring provinces characterized by distinct precipitation variability. The selection of different types of indicators and different moments of the vegetative cycle to characterize drought conditions, the complex precipitation regime, and the variable thresholds of crop-loss between provinces and cereal species, reflects how distinct the crop-specific crop-loss attributable to drought conditions can be at the sub-national scale. Previous studies performed by the authors assessing correlation patterns between SPEI and wheat yield anomalies at the province level of the IP (Páscoa et al., 2017b) have also found some inconsistencies associated to neighbouring provinces with opposite patterns. Nevertheless, in the present study the Galician provinces in the northwestern edge of the IP (Coruna, Lugo and Orense) stand out as provinces with low agricultural drought risk values in the case of both cereals in both severe or extreme and mild or moderate drought.

The present analysis might suggest that crop production in the IP could be focused on wheat given the lower levels of agricultural drought risk using the applied methodology. However, the present study does not consider an economic analysis regarding the cereal's prices, the use of water resources (irrigation) during drought periods to avoid crop-losses, the use of

machineries and fertilizers, among other factors which influence the producers' choice on crop. Quiroga et al. (2011) have assessed crop yield risk to water pressures in Spain and have found that irrigated crops do not show evidence of significant impact of drought on their yield. Although the irrigation area has an important effect in primarily rainfed crops, such as wheat, the impact decreases after a given amount of water (Quiroga et al., 2011). Therefore, some losses are still expected under non-drought conditions, as shown in the present study (Figure 4.5. and Figure 4.6). Future research using the methodology applied in the present study could differentiate the yield of wheat and barley under rainfed or irrigated conditions to address if the risk of crop-loss under drought conditions is reduced in the irrigated areas and if the costs involved in irrigation practices are compensatory. The current study considers the total area and total production given that the amount of wheat and barley in rainfed area is much higher than in irrigated area in most of the provinces, while some provinces do not use irrigation at all.

The application of this study using drought information under climate change scenarios and using seasonal drought forecasts are potential goals for future work. Moreover, heatwaves have also been responsible for cereals wheat-loss due to heat stress (Zampieri et al., 2017), and it would be important to address its influence on the likelihood of crop-loss using the presented methodology.

In summary, our results point to the fact that wheat and barley are crops vulnerable to drought conditions, whereas barley is suggested to exhibit greater probabilities of joint extreme low values between yield and drought indicators in most provinces. In such cases, drought hazard is manly characterized by VCI and SPEI, while TCI is mainly used in copula models indicating greater probabilities of joint extreme high values of wheat and drought indicators. In addition, Salamanca stands out as the province with the greatest agricultural drought risk level during mild or moderate and severe or extreme drought. From an operational point of view, the results aim to contribute to the decision-making process in agricultural practices.

Supplementary Material

Table S4.1 - Selected copula according to AIC, respective parameter (ρ in the case of Gaussian, μ in the case of t and θ in the case of Clayton, Frank or Gumbel) and the 95% confidence interval (ci).

	Provinces	Wheat				Barley			
Portugal		Copula	ρ/df/θ	ci (95%)		Copula	ρ/df/θ	ci (95%)	
	Entre Douro e Minho	Clayton	1.2	0.60	1.86	Frank	1.3	-0.53	3.18
	Trás-os-Montes	-	-	-	-	-	-	-	-
	Beira-litoral	-	-	-	-	-	-	-	-

	Beira-interior	Frank	3.7	1.04	6.30	Gaussian	0.4	0.02 0.64
	Ribatejo-oeste	Frank	-3.3	-5.66	-0.92	t	1.1	0.29 1.91
	Alentejo	Gumbel	1.8	1.27	2.32	Clayton	1.5	0.60 2.41
	Algarve	Clayton	1.0	0.08	2.02	Clayton	1.2	0.31 2.14
Spain	Coruna	Gaussian	-0.2	-0.51	0.17	Gaussian	-0.3	-0.57 0.09
	Lugo	Gumbel	1.5	0.94	2.11	Frank	-1.2	-3.23 0.74
	Orense	Frank	-3.2	-5.62	-0.78	Frank	-2.6	-5.12 -0.03
	Pontevedra	-	-	-	-	-	-	- -
	Asturias	-	-	-	-	-	-	- -
	Cantabria	-	-	-	-	-	-	- -
	Alava	Gaussian	-0.2	-0.51	0.18	Clayton	0.3	-0.25 0.83
	Guipúzcoa	-	-	-	-	-	-	- -
	Vizcaya	-	-	-	-	-	-	- -
	Navarra	Gaussian	0.5	0.12	0.70	Gaussian	0.5	0.23 0.75
	Rioja	-	-	-	-	-	-	- -
	Huesca	Frank	5.1	2.39	7.85	Gumbel	2.1	1.37 2.75
	Teruel	Frank	6.2	3.04	9.32	Gumbel	2.9	2.05 3.84
	Zaragoza	Gumbel	1.9	1.38	2.36	Gumbel	2.2	1.44 2.98
	Barcelona	Frank	7.2	3.76	10.70	Clayton	1.5	0.60 2.39
	Gerona	Clayton	1.2	0.18	2.21	Clayton	1.5	0.67 2.24
	Lerida	Gumbel	1.5	1.	1.99	Clayton	0.5	-0.64 1.68
	Tarragona	Frank	7.3	3.79	10.79	Frank	6.6	3.49 9.76
	Avila	Gaussian	0.8	0.68	0.92	Gaussian	0.8	0.55 0.88
	Burgos	t	1.4	0.13	2.58	Frank	9.1	5.45 12.75
	Leon	Gumbel	1.9	1.33	2.55	Gumbel	2.2	1.51 2.83
	Palencia	Frank	6.9	3.10	10.77	Gumbel	2.5	1.84 3.08
	Salamanca	Clayton	3.2	1.57	4.82	Clayton	2.9	1.37 4.45
	Segovia	t	1.0	0.24	1.77	Gumbel	2.5	1.61 3.46
	Soria	Gaussian	0.7	0.44	0.84	Gaussian	0.8	0.53 0.87
	Valladolid	Frank	7.7	4.37	11.06	Clayton	3.1	1.53 4.70
	Zamora	Gumbel	2.0	1.31	2.77	Clayton	1.1	0.24 1.99
	Madrid	Clayton	1.2	0.24	2.18	Clayton	0.9	0.16 1.73
	Albacete	Gumbel	1.7	1.10	2.35	Gaussian	0.9	0.71 0.93
	Ciudad-real	Gumbel	1.4	0.96	1.90	Gumbel	2.4	1.53 3.25
	Cuenca	Gaussian	0.8	0.56	0.88	Clayton	2.9	1.85 3.99
	Guadalajara	Gumbel	1.9	1.35	2.46	Gaussian	0.8	0.54 0.87
	Toledo	Clayton	2.1	0.50	3.69	Clayton	2.8	1.52 4.13
	Alicante	t	1.2	0.20	2.29	Gaussian	0.7	0.42 0.83
	Castellon	Gumbel	1.8	1.35	2.32	Clayton	0.9	0.12 1.68
	Valencia	Clayton	2.2	0.68	3.69	Gaussian	0.8	0.57 0.88
	Murcia	Gumbel	1.7	1.12	2.36	Clayton	1.6	0.68 2.61
	Badajoz	Clayton	1.1	0.03	2.10	Clayton	1.8	0.75 2.81
	Caceres	Gumbel	1.5	1.03	1.88	Clayton	1.4	0.41 2.42
	Almeria	Clayton	2.1	1.10	3.04	Clayton	1.9	0.73 2.98
	Cadiz	Clayton	1.0	0.29	1.71	Gumbel	1.7	1.24 2.13
	Cordoba	Clayton	1.8	0.63	3.	Clayton	1.5	0.58 2.49
	Granada	Gaussian	0.6	0.29	0.78	Frank	9.1	5.34 12.83
	Huelva	Frank	7.4	3.88	10.88	Frank	5.8	2.81 8.79
	Jaen	Clayton	1.8	0.82	2.80	Clayton	2.1	1.14 3.06
	Malaga	Clayton	1.2	0.27	2.19	Clayton	1.2	0.25 2.07
	Sevilla	Clayton	1.6	0.86	2.42	Clayton	1.9	0.64 3.07

Chapter 5

Risk of crop failure due to compound dry and hot extremes estimated with nested copulas

***Ribeiro, A.F.S.**, Russo, A., Gouveia, C.M., Páscoa, P., Zscheischler, J. (2020). Risk of crop failure due to compound dry and hot extremes estimated with nested copulas, Biogeosciences Discussions, https://doi.org/10.5194/bg-2020-116*

***Andreia F. S. Ribeiro contributions:** preparation of the climate and crop data; statistical analysis and interpretation of the results (nested Archimedean copulas and estimation of conditional probabilities); production of all the figures, coordination and writing of the manuscript.*

Graphical abstract

Data

Weather data

Spring maximum temperature ($Tmax_{MAM}$)
Spring precipitation (P_{MAM})

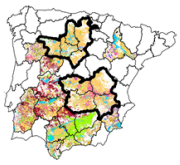
Compound dry and hot extremes

	Dry	Hot
Moderate (+)	$P_{MAM} \leq 20^{th}$	$Tmax_{MAM} \geq 80^{th}$
Severe (++)	$P_{MAM} \leq 10^{th}$	$Tmax_{MAM} \geq 90^{th}$
Extreme (+++)	$P_{MAM} \leq 5^{th}$	$Tmax_{MAM} \geq 95^{th}$

Crop data

Wheat and barley t/ha

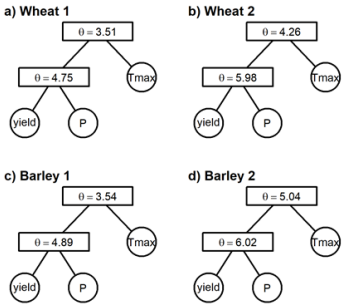
Threshold for crop failure
20^{th} percentile



Methods

Nested Archimedean Copulas (NAC)

$$C(u_1, u_2, u_3) = C_1(C_{12}(u_1, u_2; \theta_{12}), u_3; \theta_1)$$



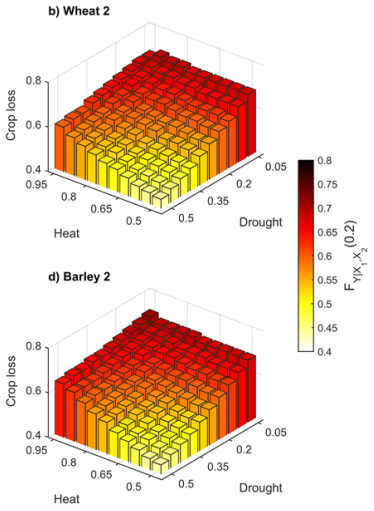
Agricultural impacts

Under dry conditions
$F_{Y|X_1}(Y|X_1 = x_1^*) = P(Y \leq y|X_1 \leq x_1^*)$

Under hot conditions
$F_{Y|X_2}(Y|X_2 = x_2^*) = P(Y \leq y|X_2 \leq x_2^*)$

Under compound dry and hot conditions
$F_{Y|X_1,X_2}(Y|X_1 = x_1^*, X_2 = x_2^*) = P(Y \leq y|X_1 \leq x_1^*, X_2 \leq x_2^*)$

Results



Crop loss due to compound dry and hot conditions is driven primarily by **drought** than by heat

Highlights

- A dependence between crop yields, drought and heat conditions is identified based on Nested Archimedean Copulas (NAC)
- Agricultural impacts aggravate during concurrent hot and dry conditions, when compared with the crop-loss induced by individual hot or dry conditions
- The probability of crop-loss increases with the severity of the compound hot and dry extremes in both regions (northern and southern) and cereals (wheat and barley)
- Drought plays the dominant role in the compound event
- Higher chances of crop-loss occur in the southern region of the Iberian Peninsula, particularly in the case of barley

Abstract

The interaction between co-occurring drought and hot conditions is often particularly damaging to crop's health and may cause crop failure, which can threaten food security. Climate change exacerbates such risks due to an increase in the intensity and frequency of dry and hot events in many land regions. Here we model the trivariate dependence between spring maximum temperature, spring precipitation and wheat and barley yields, respectively, over two province clusters in Spain with nested copulas. Based on the full trivariate joint distribution, we (i) estimate the impact of compound hot and dry conditions on wheat- and barley-loss and (ii) estimate the additional impact due to compound hazards compared to individual hazards. We find that crop-loss increases when drought or heat-stress aggravates to compound dry and hot conditions and that an increase in the severity of compound conditions leads to larger damages. For instance, compared to moderate drought only, compound dry and hot conditions increase the likelihood of crop-loss by 8 to 11% while when starting with moderate heat, the increase is between 19 to 29% (depending on the cereal and region). These findings suggest that the likelihood of crop-loss is driven primarily by drought stress than by heat stress, suggesting that drought plays the dominant role in the compound event, that is, drought stress does not require to be so extreme as heat stress to cause a similar damage. Furthermore, when compound dry and hot conditions aggravate from moderate to severe or extreme stress, crop-loss probabilities increase 5 to 6% and 6 to 8% respectively (depending on the cereal and region). Our results highlight the additional value of a trivariate approach for the estimating the compounding effects of dry and hot extremes on of crop failure risk. Therefore, this approach can effectively contribute to design management options and guide the decision-making process in agricultural practices.

Keywords: Compound events; Agricultural impacts; Dry and hot conditions; Trivariate dependence; Nested copulas

5. Risk of crop failure due to compound dry and hot extremes estimated with nested copulas

5.1 Introduction

The assessment of the adverse social, economic and environmental impacts associated with a combination of multiple climate hazards have recently become a focus of high interest (Leonard et al., 2014; Zscheischler et al., 2020). Such compound events often lead to larger impacts compared to when hazards occur separately (Zscheischler et al., 2018). For instance, compound dry and hot conditions reduce carbon uptake more strongly compared to the sum of the individual hazards (Zscheischler et al., 2014).

Dry and hot conditions often co-occur. For instance in Europe, the extreme 2003, 2010 and 2018 heatwaves were accompanied by strong soil moisture deficits (Bastos et al., 2014; Schumacher et al., 2019). In 2010, the compound event was particularly strong in Russia (Schumacher et al., 2019), while in 2003 the extreme drought and heatwave affected mostly central Europe, extending to west Mediterranean countries like Portugal and Spain (Garcia-Herrera et al., 2010), with critical consequences in several sectors. In 2010, widespread crop yield declines and failures occurred over the major grain producing regions of Russia, northeastern Ukraine, and northwestern Kazakhstan (Loboda et al., 2017). Previously, the shortages in crop yields in 2003 have also caused major financial losses in the agricultural sector, and when compared to the previous year, the cereal productions in European Union (EU) have decreased 23 million tonnes (COPA-COGECA, 2003). The decline in the harvests was both in quantity and quality, as was the case of winter cereals whose maturation was accelerated due to compound extreme dry and hot conditions, forming grains with insufficient water content (COPA-COGECA, 2003). The 2018 event strongly impacted pastures and arable land north of the Alps (Buras et al., 2020). As the occurrence of climate extremes such as heatwaves, droughts and compound dry and hot events is expected to increase in intensity and frequency in many land regions due to climate change (IPCC, 2012; Zscheischler and Seneviratne, 2017), associated adverse impacts such as widespread harvest failures threatening global cereals supplies may also increase.

Among the panoply of multivariate approaches applied to assess the impacts of multiple climate hazards, the use of copulas has become quite popular in studies focused on analysing the social, environmental and economic risks associated with adverse climate conditions (Bokusheva et al., 2016; Madadgar et al., 2017; Zscheischler and Seneviratne, 2017; Gaupp et al., 2019; Ribeiro et al., 2019c, 2019a). With copulas nonlinear dependency structure can be modelled, which offers more flexibility and possibly a more adequate fit for different dependence types in the extremes. (Nelsen, 2006; Salvadori and De Michele, 2007; Durante and Sempi, 2015). Among all types of copulas described in the literature, the popularity of the class of Elliptical copulas comes from the fact that they derive from well known distributions associated to the widely used Pearson's correlation, but the elliptical dependence is only able to capture radial symmetry and the respective mathematical expressions do not have a closed form. One of the copula classes that overcomes this drawback is the Archimedean, which have a simpler mathematical form and can capture different kinds of tail dependence and radial symmetry or asymmetry.

Archimedean copulas (AC) are exchangeable, which means that the copula is the same if we permute the respective margins. For the bivariate case this may not be a limitation, but as the number of dimensions increase, it is unlikely that exchanging across the involved variables allows for the 'true' dependence structure to be well-defined. To avoid exchangeability, nested Archimedean copulas (NAC) have been proposed (Okhrin and Ristig, 2014), also referred to as hierarchical Archimedean copulas (HAC), obtained by nesting lower dimensional Archimedean copulas into each other and/or with marginal distributions. Okhrin and Ristig (2014) introduced NACs where all copulas belong to the same family with a nesting condition that requires decreasing dependence strength from the highest to the lowest hierarchical level. Here we make use of this NAC approach, taking advantage of the balance between flexibility (modelling different types of dependence structures) and usability in higher dimensions (limiting the number of parameters).

The present work aims to identify how risks associated with compound dry and hot conditions affect wheat and barley yields over two clusters of provinces in Spain based on the trivariate dependence between precipitation, maximum temperature and yields using a NAC approach. In particular, we are interested in quantifying the additional risk associated with compound dry and hot conditions compared to only dry or only hot conditions. Wheat and barley are chosen as they are two of the major rainfed crops in the Iberian Peninsula (Vicente-Serrano et al., 2006;

Peña-Gallardo et al., 2019). Moreover, we here build on prior work which has estimated wheat- and barley-losses in the same area, but related to a single hazard, namely droughts (Ribeiro et al., 2019b, 2019c).

Using NACs, we estimate the conditional probabilities of crop-loss under different severity levels of dry and hot conditions based on the full trivariate joint distribution. We focus on annual wheat and barley yield data at the sub-national scale, thus overcoming drawbacks related to assessing climate related crop risks at the national scale. Based on the proposed approach we (i) characterize the dependence structures between the dry and hot conditions and the crop yields; (ii) estimate the conditional probability of crop-loss under different compound dry and hot severity levels; and (iii) evaluate how much the compound dry and hot conditions increase the risk of crop failure in comparison to the individual hazards.

5.2 Data and Methods

5.2.1 Crop yield data

Wheat and barley yields were obtained for 9 provinces in Spain from the Spanish Agriculture, Fishing and Environment Ministry (available at https://www.mapa.gob.es/es/estadistica /temas/publicaciones/anuario-de-estadistica/, last access on 9 November 2019). Those 9 provinces were assembled in two distinct clusters (Figure 5.1) which are dominated by rainfed agricultural practices considering the non-irrigated arable land classification from CORINE Land Cover dataset based on an earlier clustering (Ribeiro et al., 2019b, 2019c). The provincial clustering consisted in the application of three main criteria: first the provinces with land use dominated by agricultural practices were identified (Figure 5.1), and from those provinces, the ones dominated by non-irrigated practices and contiguous in space were selected for analysis (Figure 5.1 - bold black contours). Figure 5.1 shows the Iberian provinces with < 50% agricultural pixels colored in white, the provinces with > 50% agricultural pixels colored with the respective agricultural CORINE classes and the selected two clusters of contiguous provinces dominated by rainfed agriculture delineated in bold black contours. This aggregation of provinces allowed for the identification of two major cereals' production areas which are dominated by rainfed cropping systems among the provinces with higher percentage of agricultural land use in the IP.

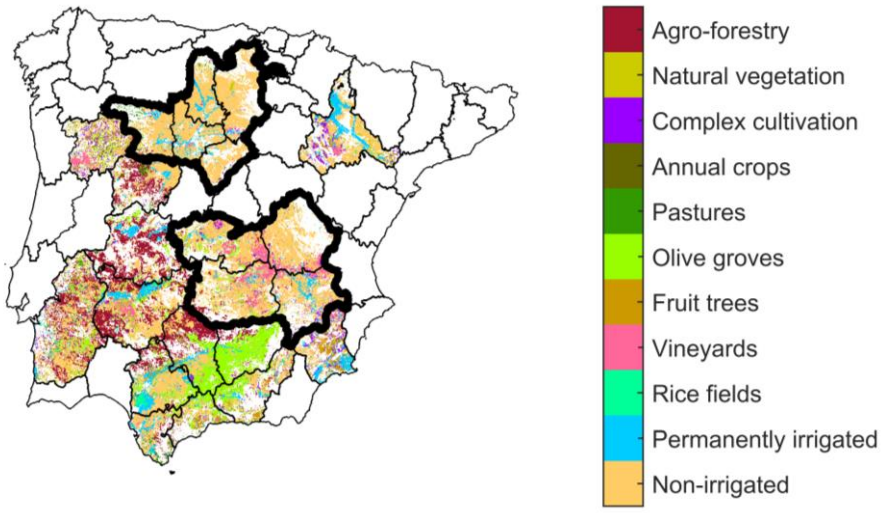


Figure 5.1 - Iberian provinces dominated by agricultural land use (> 50% agricultural pixels belonging to all agricultural CORINE classes, see legend) according to the CORINE Land Cover dataset and respective categories. The contiguous provinces dominated by rainfed practices (> 50% non-irrigated pixels in yellow) are delineated in bold black contours and grouped in two clusters. Northern region (cluster 1) provinces: Burgos, Palencia, Segovia, Valladolid, and Zamora. Southern region (cluster 2) provinces: Albacete, Ciudad Real, Cuenca, and Toledo.

Crop yields were obtained as the ratio between production and harvested area during the period of 1986–2016.We computed crop yield anomalies by removing longer term trends based on locally estimated scatterplot smoothing (LOESS, a method for local regression) to account for yield increases due to technological development (Ben-Ari et al., 2016). We pooled crop yields from the provinces over each cluster, resulting in samples sizes $N_1 = 155$ for cluster 1 (30 years of annual data over five provinces) and $N_2 = 124$ for cluster 2 (30 years of annual data over four provinces). Pooling time series greatly expands the sample size allowing greater robustness in three-dimensional statistical analysis that otherwise would be compromised. This type of assessment is a compromise between the use of a sub-national resolution of crop data and the sample size to evaluate the number of cases of simultaneous occurrence of dry and hot conditions.

5.2.2 Weather data

The vegetative cycle of the winter crops in Spain is mainly driven by precipitation and temperature: sowing occurs around autumn, followed by the vegetative phase in winter, reproductive phase in spring (when vegetation is photo-synthetically more active) and crop harvest occurs in the early summer. Therefore, monthly precipitation (P) and monthly maximum temperature (Tmax) were extracted from the Climate Research Unit (CRU) TS4.01 dataset (Harris et al., 2014) spanning the same time period. We used 3-monthly means of Tmax

and 3-monthly means of P during spring (P_{MAM} and $Tmax_{MAM}$, respectively), which was identified as the most sensitive time period for crop yield based on correlation analysis (Figure 5.2). This selection of climate variables allows to maximize the dependence between climate conditions and yields as also shown by previous work based on the same data (Ribeiro et al., 2019b).

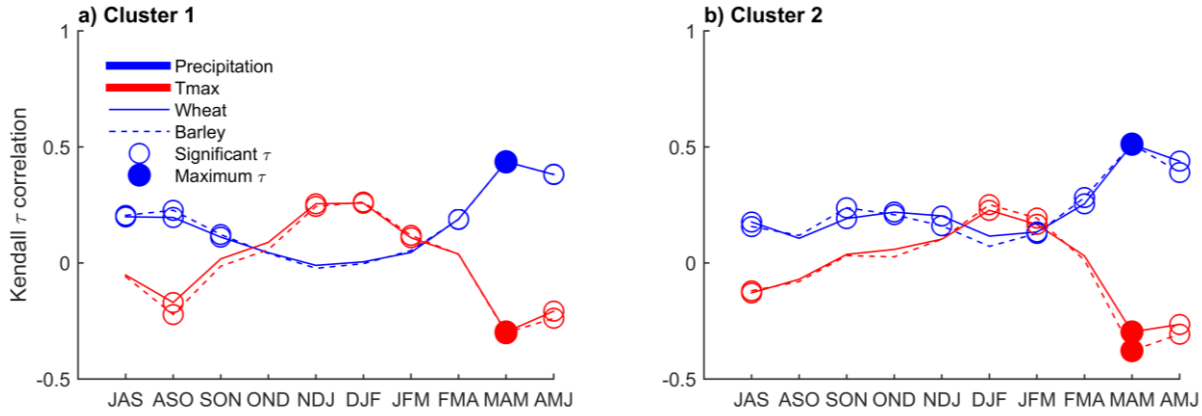


Figure 5.2 - Kendall correlation τ between three-monthly means of maximum temperature (Tmax, red) and precipitation (blue) with wheat (filled lines) and barley (dashed lines) yield anomalies, respectively. Correlations were computed during the crop growing period (September to June) over 1986-2016 for cluster 1 (a) and 2 (b) (Figure 5.1). The months on the x-axis denote the end-month of the averaging period. Circles indicate statistically significant correlations at $\alpha = 0.05$. The strongest correlation (positive or negative) is denoted by filled circles (P_{MAM} and $Tmax_{MAM}$).

We considered three severity levels of dry and/or hot conditions: Moderate (+), Severe (++) and Extreme (+++) based on percentile thresholds as shown in Table 5.1. Besides these three severity levels, we further considered all combinations of 10 categories of severity levels of dry and hot conditions exceeding the 50^{th} to 5^{th} and 50^{th} to 95^{th} percentiles for P_{MAM} and $Tmax_{MAM}$, respectively. We further considered the 20^{th} percentile of the crop anomaly time-series as lower exceedance threshold for crop failure (Ben-Ari et al., 2016; Ribeiro et al., 2019a, 2019c).

Table 5.1 - Categories of severity levels of dry and hot conditions based on P_{MAM} and $Tmax_{MAM}$ percentiles.

	Moderate (+)	Severe (++)	Extreme (+++)
dry	$P_{MAM} \leq$ 20th	$P_{MAM} \leq$ 10th	$P_{MAM} \leq$ 5th
hot	$Tmax_{MAM} \geq$ 80th	$Tmax_{MAM} \geq$ 90th	$Tmax_{MAM} \geq$ 95th

5.2.3 Modelling trivariate distributions with nested Archimedean copulas

We model the trivariate relationship between temperature, precipitation and crop yields with nested copulas. Consider a vector of crop yield annual anomalies Y and the climate variables $X_1 = P_{MAM}$ and $X_2 = Tmax_{MAM}$ with marginal cumulative distribution functions (CDF) F_Y, F_{X_1} and F_{X_2}, respectively. We aim to estimate and compare three conditional cumulative distribution functions (CDFs) with the scalars x_1^* and x_2^* corresponding to the dry and hot thresholds, respectively:

$$F_{Y|X_1}(Y|X_1 = x_1^*) = P(Y \leq y|X_1 \leq x_1^*) \quad (5.1)$$

$$F_{Y|X_2}(Y|X_2 = x_2^*) = P(Y \leq y|X_2 \leq x_2^*) \quad (5.2)$$

$$F_{Y|X_1,X_2}(Y|X_1 = x_1^*, X_2 = x_2^*) = P(Y \leq y|X_1 \leq x_1^*, X_2 \leq x_2^*) \quad (5.3)$$

With the above equations we can estimate the agricultural impacts under dry conditions $F_{Y|X_1}$ (Equation 5.1), under hot conditions $F_{Y|X_2}$ (Equation 5.2) and under compound dry and hot condition $F_{Y|X_1,X_2}$ (Equation 5.3), respectively. In other words, if the compound dry and hot conditions cause more damage than the individual hazards, it is expected that $F_{Y|X_1,X_2}$ suggests higher probabilities of crop-loss (i.e., y = y*for a low y) than $F_{Y|X_1}$ or $F_{Y|X_2}$. Furthermore, we can study the relative role of P_{MAM} and $Tmax_{MAM}$ in crop-loss with Equation 5.1 and Equation 5.2.

To compare the additional impact of compound dry and hot conditions with the impacts caused by the individual hazards, Equations 5.1-5.3 are used to estimate

$$Relative\ change\ from\ drought - stress = \frac{F_{Y|X_1=x_1^*,X_2=x_2^*}(0.2) - F_{Y|X_1=x_1^*}(0.2)}{F_{Y|X_1=x_1^*}(0.2)} \quad (5.4)$$

$$Relative\ change\ from\ heat - stress = \frac{F_{Y|X_1=x_1^*,X_2=x_2^*}(0.2) - F_{Y|X_2=x_2^*}(0.2)}{F_{Y|X_2=x_2^*}(0.2)} \quad (5.5)$$

where 0.2 is the threshold of crop-loss (y*) corresponding to the 20th percentile of the crop yields. These changes can be estimated for different severity levels of dry (x_1^*) and hot (x_2^*) conditions.

Following the theorem of Sklar (1959) we can decompose a multivariate probability distribution into its marginals and a copula C which describes the dependence structure between the margins. To estimate the multivariate distribution $P(Y, X_1, X_2)$, the respective copula C is fitted, which is then a joint CDF whose marginal distributions are uniform in the interval [0; 1] (Nelsen, 2006; Salvadori and De Michele, 2007; Durante and Sempi, 2015). Transforming the margins to uniform variables through their CDFs, that is, $u_1 = F_Y$, $u_2 = F_{X_1}$ and $u_3 = F_{X_2}$, the trivariate CDF can be written as (Sklar, 1959):

$$F(u_1, u_2, u_3) = C(u_1, u_2, u_3) \tag{5.6}$$

Within the copula families, AC are extensively used due to their flexibility and applicability to a variety of tail dependence structures, as well as their analytical tractability. AC can be written in terms of the respective generator function φ, e.g. for the three-dimensional case:

$$C(u_1, u_2, u_3; \theta) = \varphi_\theta(\varphi_\theta^{-1}(u_1) + \varphi_\theta^{-1}(u_2) + \varphi_\theta^{-1}(u_3)) \tag{5.7}$$

Due to the symmetry of bivariate AC, the above trivariate form can be expressed in terms of NAC or HAC, where two of the margins are coupled by their bivariate copula,

$$C(u_1, u_2, u_3; \theta_{12}; \theta_1) = C_1(C_{12}(u_1, u_2; \theta_{12}), u_3; \theta_1) \tag{5.8}$$

Equation 5.8 can also be expressed in terms of the other possible pair copulas $C_{13}(u_1, u_3; \theta_{13})$ and $C_{23}(u_2, u_3; \theta_{23})$ that are coupled with u_2 and u_1 by C_2 and C_3 , with expressions $C_2(C_{13}(u_1, u_3; \theta_{13}), u_2; \theta_2)$ and $C_3(C_{23}(u_2, u_3; \theta_{23}), u_1; \theta_3)$, respectively.

Most structures of NAC require decreasing parameters from the inner to the outer hierarchical level to attain a properly fitted copula. As for most ACs, the larger the parameter the stronger the dependence, this means that most structures of NAC require that the marginal copulas in the inner level should correspond to the pair with the strongest dependence, i.e., satisfying $\theta_{12} \geq \theta_1$ in the case of Equation 5.8. This requirement applies to NAC with generators from the same family, providing a flexible estimation of the NAC, which allows for specifying the full distribution with at most d-1 parameters, where d is the number of copula dimensions or marginal distributions (Okhrin and Ristig, 2014).

In our study we focus on a total of four Archimedean families that capture different kinds of joint dependence structures: Clayton, Gumbel, Frank and Joe. The Clayton, Gumbel and Joe

copulas describe an asymmetrical tail behaviour, while the Frank copula, in a similar way to the Gaussian copula, captures joint symmetric dependence. While Gumbel and Joe copulas can represent upper tail dependence, Clayton copulas can represent lower tail dependence. The estimation of the copula parameters is based on maximum likelihood based on the R package HAC (Okhrin and Ristig, 2014).

The main steps of the trivariate approach used in this study can be summarized as follows (Okhrin and Ristig, 2014). First, the marginal distributions u_1, u_2 and u_3 are estimated non-parametrically by simple ranking, a common approach for copula modelling. Afterwards, the fit of bivariate copula models is performed to every pair of variables to estimate C_{12} , C_{13} and C_{23}. For each pair, the copula selection is performed based on the Akaike's information criterion (AIC) and checking the goodness-of-fit by comparing the empirical copula based on the Cramer-von Mises distance (Sn). The bivariate copula with the strongest dependence, with the lowest AIC and the lowest Sn, is selected to define the structure of the NAC. Afterwards, the marginal distribution that is not part of the selected bivariate copula is joined and the parameter of the upper level copula of the same family is estimated (Equation 5.8). As a final step, the estimated NAC with two parameters is compared with the same Archimedean family with one parameter (Equation 5.7) in terms of the AIC, which penalizes the number of estimated parameters.

5.2.4 Diagnostics and uncertainties in the estimation procedure

The visual diagnostics of the quality of the selected models are performed analogously to a QQ-plot by comparing the empirical estimate of the Kendall function (cumulative distribution of the copula) with the theoretical estimate of the Kendall function based on the selected parametric trivariate copulas (Okhrin and Ristig, 2014).

Best estimates of all conditional probabilities (i.e., Equations 5.1-5.5) are estimated by drawing N = 100,000 samples from the fitted trivariate copula. Due to the negative dependence between $Tmax_{MAM}$ and both crop yields and P_{MAM}, we inverted the margins of $Tmax_{MAM}$ for copula modelling.

Uncertainties of the statistical modelling are estimated by repeated sampling (10,000 times) of the fitted model with sample sizes equal to the number of observations (i.e., N_1 in the case of

cluster 1 and N_2 in the case of cluster 2). From these samples, 95% confidence intervals of Kendall's rank correlation are estimated and compared with the observed pairs (u1,u2), (u1,u3) and (u2,u3). This validation step intends to verify if the generated pairs of copula-based samples preserve the level of dependence found in the observations. Furthermore, this approach is used to estimate uncertainties related to the conditional probabilities (Equations 5.1-5.5).

5.3 Results

In both cereals and both clusters the most dependent pair of variables corresponds to crop yields and P_{MAM}, hence the pair of variables u1,u2 defines the optimal NAC structure (Figure 5.3). Results for all possible variable pairs and the respective bivariate copulas are shown in Table S5.1.

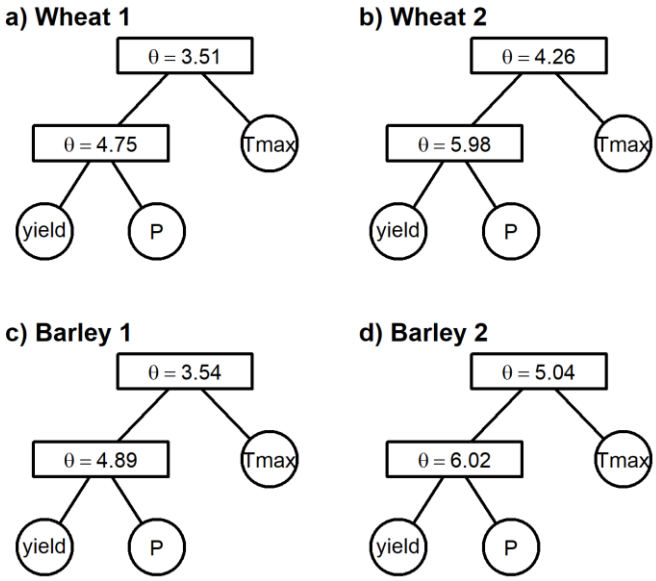


Figure 5.3 - Structure and respective parameters of the selected nested Frank models $C_1(C_{12}(u_1, u_2; \theta_{12}), u_3; \theta_1)$ to model the trivariate joint distributions between crop yields, P_{MAM} and $Tmax_{MAM}$. (a) Wheat in cluster 1. (b) Wheat in cluster 2. (c) Barley in cluster 1. (d) Barley in cluster 2.

Once the bivariate copula $C_{12}(u_1, u_2)$ of yields and P_{MAM} is known, the NAC models are constructed (Table 5.2). The Frank copula provides the best fit of $C_{12}(u_1, u_2)$ (Table S5.1) for both cereals and both clusters and thus the parameters of the trivariate nested copulas are all from the Frank family. Nevertheless, despite Frank being the best family to characterize the nested copulas, we also constructed NAC models with Gumbel, Clayton and Joe copulas for

comparison, as well as trivariate Archimedean copulas with one parameter where we selected the best structure between one-parameter and two-parameter AC copulas via the AIC (Table 5.2). In all but one case the NAC models with Frank copulas is the best model. The only exception is barley in cluster 2 whose AIC of $C_\theta(u_1, u_2, u_3)$ is slightly lower than the AIC of $C_{\theta_1}\left(u_3, C_{\theta_{12}}(u_1, u_2)\right)$ (Table 5.2). Nevertheless, in terms of Cramer-von Mises distance (Sn) the nested copula is the closer to the empirical trivariate copula. For this reason, we modelled the trivariate joint distribution based on nested Frank copulas for all cases. For all fitted models, the empirical cumulative distribution corresponds well to the theoretical cumulative distributions (Figure 5.4). Bivariate dependencies as measured by Kendall's τ are captured well by the fitted models (Figure 5.5 for wheat, Figure S5.1 for barley).

Table 5.2 - Trivariate Archimedean copulas (AC) parameters $\boldsymbol{\theta}$ with nested structure with two-parameters (C(C(u1,u2),u3) and with one-parameter (C(u1,u2,u3)) and respective Akaike's Information Criteria (AIC) and Cramer-von Mises distance (Sn). Fit based on maximum pseudo-likelihood (Gumbel (G), Clayton (C), Frank (F) and Joe (J) copulas). Smaller values of AIC and Sn indicate the selected copula for each cereal and cluster (bold and underlined).

			Cluster 1					**Cluster 2**			
			G	**C**	**F**	**J**		**G**	**C**	**F**	**J**
Wheat	$C_\theta(u_1, u_2, u_3)$	θ	1.41	0.66	3.22	1.53	θ	1.53	0.75	3.88	1.72
		AIC	-74.16	-79.89	**-99.02**	-49.16	AIC	-89.12	-74.67	**-106.14**	-69
		Sn	0.15	0.21	**0.07**	0.31	Sn	0.14	0.31	**0.07**	0.27
	$C_{\theta_1}(C_{\theta_{12}}(u_1, u_2), u_3)$	θ_1	1.37	0.9	3.51	1.41	θ	1.57	0.91	4.26	1.76
		θ_{12}	1.59	0.93	4.75	1.73	θ	1.88	1.37	5.98	2.11
		AIC	-79.69	-71.27	***-102.84***	-54.29	AIC	-99.7	-79.76	***-112.93***	-78.49
		Sn	0.12	0.11	***0.03***	0.3	Sn	0.08	0.18	***0.03***	0.19
Barley	$C_\theta(u_1, u_2, u_3)$	θ	1.43	0.66	3.25	1.57	θ	1.58	0.81	4.12	1.8
		AIC	-80.8	-78.91	**-101.84**	-57.51	AIC	-105.59	-85.87	**-118.55**	-83.54
		Sn	0.12	0.21	**0.07**	0.26	Sn	0.16	0.36	**0.08**	0.3
	$C_{\theta_1}(C_{\theta_{12}}(u_1, u_2), u_3)$	θ_1	1.38	0.87	3.54	1.43	θ	1.72	1.05	5.04	1.99
		θ_{12}	1.7	0.92	4.89	1.92		1.94	1.41	6.02	2.21
		AIC	-95.8	-72.07	***-107.17***	-73.98	AIC	-112.52	-86.85	**-116.31**	-90.86
		Sn	0.09	0.12	***0.04***	0.22	Sn	0.08	0.21	***0.03***	0.19

Figure 5.4 - Empirical versus theoretical probability distributions based on the nested Frank copula models. (a) Wheat in cluster 1. (b) Wheat in cluster 2. (c) Barley in cluster 1. (d) Barley in cluster 2.

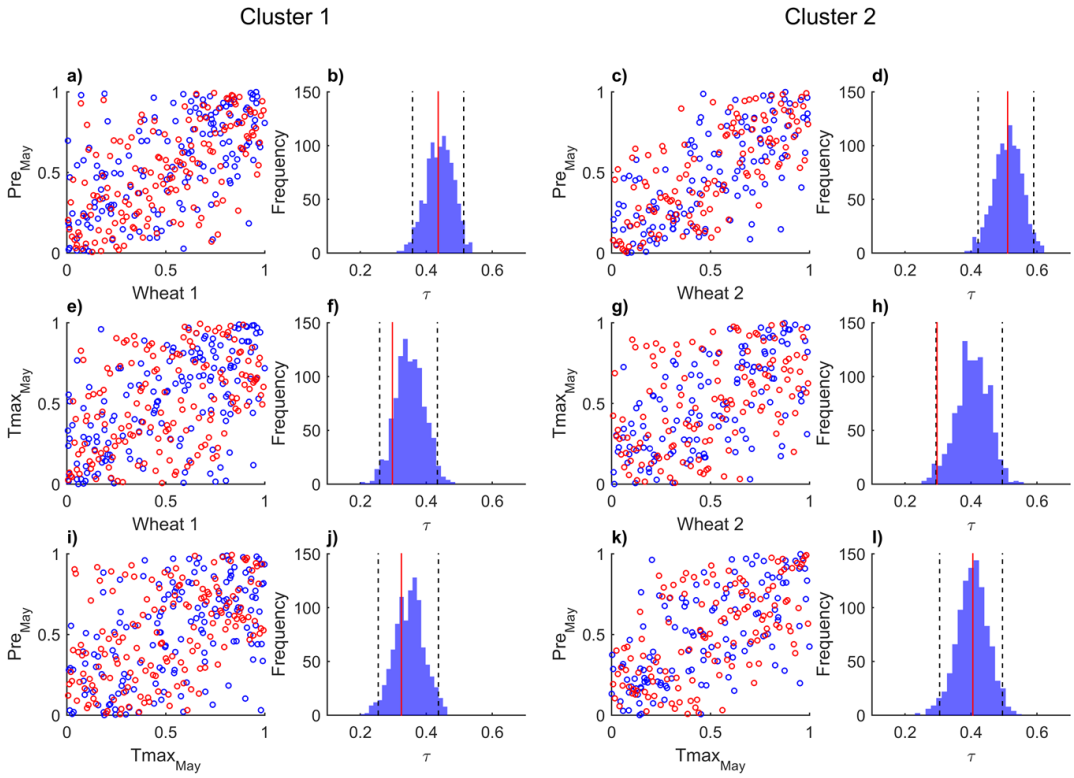


Figure 5.5 - Scatterplots of copula-based samples (blue) compared with ranked observations (red) of crop anomalies with climate variables (P_{MAM} and $Tmax_{MAM}$) (a), c), e) and g)) and P_{MAM} against $Tmax_{MAM}$ (i) and k)), for both clusters. The histograms (b), d), f), h), j), l)) correspond to the Kendall rank correlation of each pair based on 10,000 simulations with the same sample size of the observational sample. The 95% confidence intervals are shown with dashed lines. The red lines indicate the Kendall rank correlation of the observations.

The cumulative conditional probabilities of yield under moderate (+), severe (++) and extreme (+++) compound dry and hot conditions demonstrate that the probability of crop-loss increases with the severity of compound dry and hot conditions for both clusters and both cereals (Figure 5.6a-d). Moreover, the likelihood of crop-loss is higher in cluster 2 for both cereals, particularly in the case of barley. Under extreme dry and hot conditions (+++dry+++hot, purple), the likelihood of crop-loss is 68% and 71% for wheat and barley, respectively, in cluster 2, in contrast to 62% and 63% in cluster 1 (Figure 5.6e, purple bars). In addition, the differences in crop-loss are higher between moderate (+dry+hot) and severe (++dry++hot) conditions compared to the differences between severe and extreme (+++dry+++hot) conditions. More precisely, when the compound dry and hot conditions aggravate from moderate to severe stress, crop-loss increases 5 to 6% and when the compound dry and hot conditions aggravate from moderate to extreme stress, crop-loss increases 6 to 8% (depending on the cereal and region). For comparison, conditional cumulative probability distributions for single stressors compared with the compound stressors are shown in Figure S5.2 for all three severity levels.

While Figure 5.6 illustrates the same severity levels for the different hazards, Figure 5.7 illustrates crop-loss for a range of different combinations of severity levels of dry and hot conditions (e.g., extreme dry conditions combined with moderate, severe and extreme hot conditions, and vice-versa) starting from the 50th percentile of P_{MAM} and $Tmax_{MAM}$. When P_{MAM}/$Tmax_{MAM}$ are below/above the median, the probability of crop-loss is always higher than 40%. Similarly, to Figure 5.6, the increase of crop-loss with the severity of drought- and heat-stress is evident (Figure 5.7). The higher likelihood of crop-loss in cluster 2, particularly for barley, is also consistent with Figure 5.6. Moreover, the results indicate that droughts are typically associated with higher probabilities of crop-loss than heatwaves at the same severity level. This finding suggests that drought stress causes more damage to crop yields than heat stress, even for lower values of stress.

In all cases, the additional effect of compound dry and hot conditions is larger when starting from only hot conditions, compared to when starting from only dry conditions (Figure 5.8 for moderate stress, Figure S5.3a and b for severe and extreme stress). The estimates are based on Equation 5.4 and Equation 5.5. Depending on the cereal and region, the difference from drought stress to compound conditions may vary from 8% (barley in cluster 1) to 11% (barley in cluster 2). In contrast, the difference from heat stress to compound conditions may vary between 19% (barley in cluster 2) to 29% (wheat in cluster 2). Uncertainties are large for these estimates and

increase with the severity of the events (Figure S5.3). Consistent with Figure 5.7 these findings suggest that drought stress is the major driver of crop-loss associated with compound drought and heat.

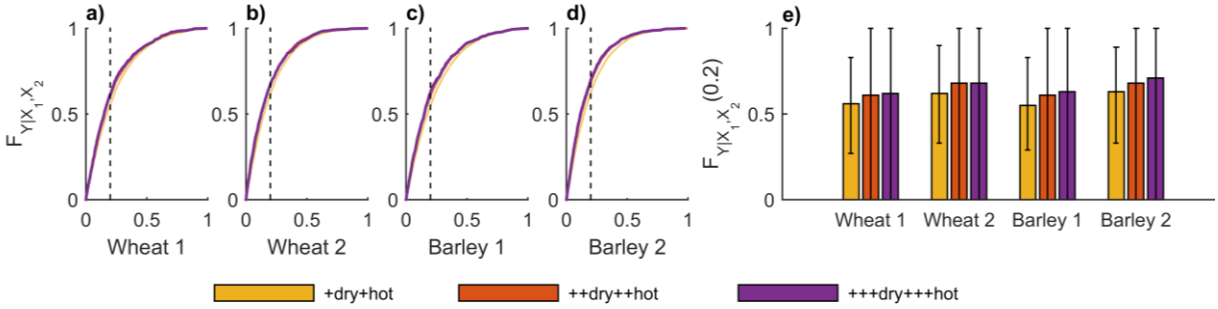


Figure 5.6 - Conditional probability distributions of crop yield anomalies $F_{Y|X_1,X_2}$ over each cluster of provinces (wheat in cluster 1 (a), wheat in cluster 2 (b), barley in cluster (1) and barley in cluster 2 (d)) under moderate (+dry+hot (yellow)), severe (++dry++hot (orange)) and extreme (+++dry+++hot (purple)) compound dry and hot conditions (see Table 5.1). (e) Conditional probabilities of non-exceeding the crop-loss threshold (20th percentile - vertical black dashed line in a) - d)) for each severity level of compound hot and dry conditions given by $F_{Y|X_1,X_2}$ (0.2). Uncertainty ranges illustrate the 95% confidence intervals.

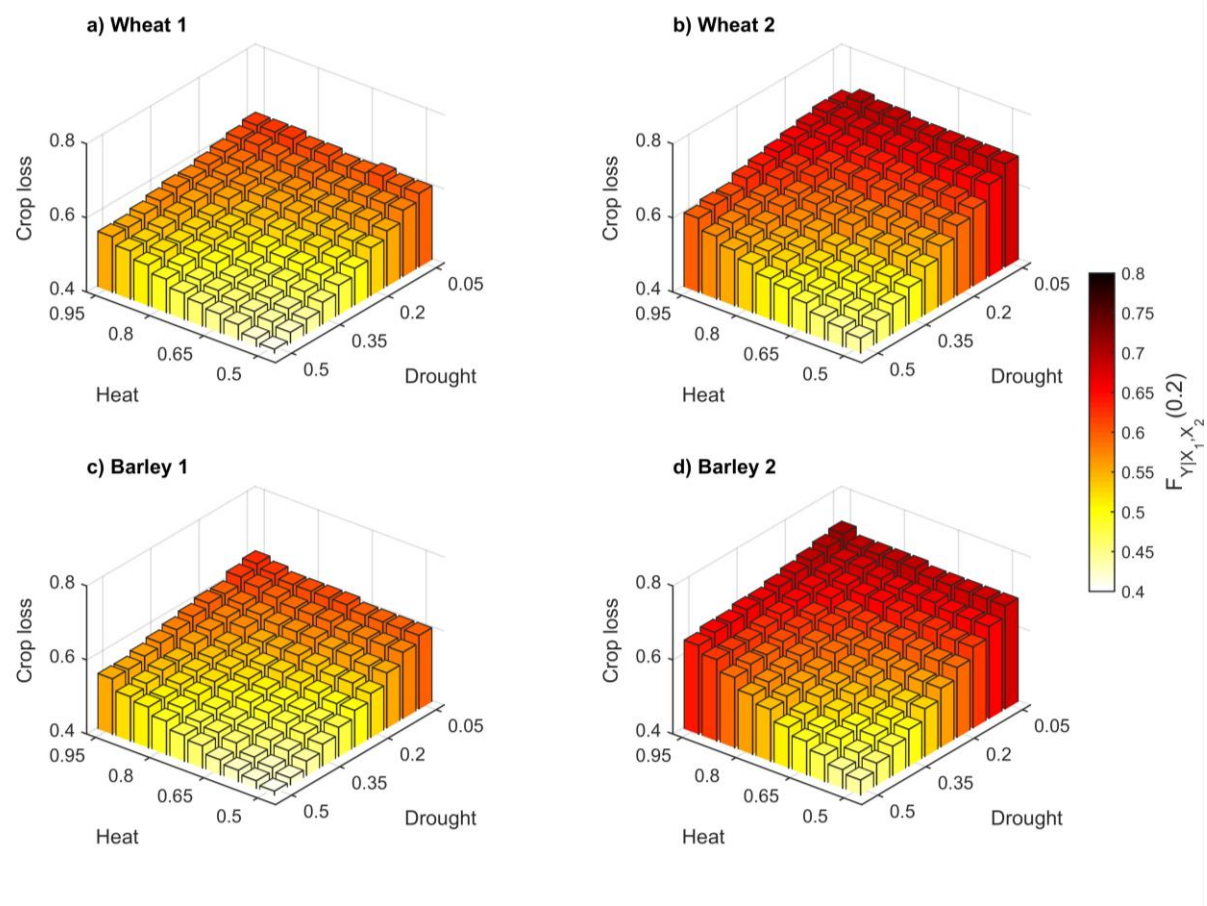


Figure 5.7 - Conditional probability of crop-loss given by $F_{Y|X_1,X_2}$ (0.2) over each cluster of provinces (wheat in cluster 1 (a), wheat in cluster 2 (b), barley in cluster 1 (c) and barley in cluster 2 (d)) for different combinations of severity levels of dry and hot conditions.

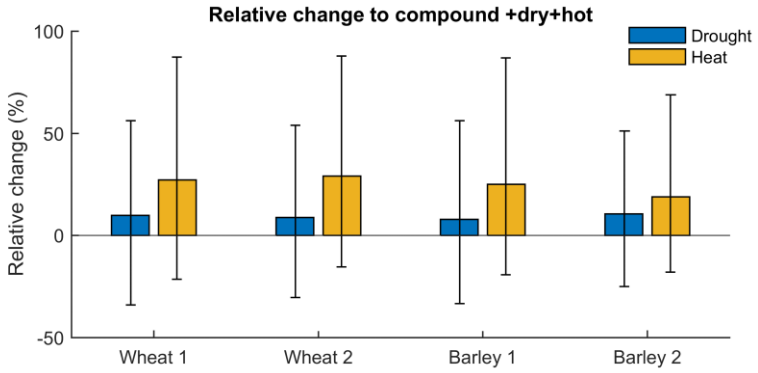


Figure 5.8 - Relative changes in crop-loss from dry (blue) or hot (yellow) influence to compound hot and dry condition at the moderate severity level (+dry+hot). Figure S5.3 show the same information for the severe (++dry++hot) and extreme (+++dry+++hot) severity levels.

5.4 Discussion

We have modelled the trivariate relationship between spring maximum temperature, spring precipitation and wheat and barley yields in two Spanish regions using nested copulas. The likelihood of crop-loss was found to increase with the severity of the compound dry and hot conditions, which is exacerbated comparatively to the occurrence of single drought or heat events. Moreover, our findings suggest that drought stress does not require to be so extreme as heat stress to cause the same adverse impact on crop yields. Hence drought is the more stressful driver of crop-loss, when considering compound drought and heat.

Although the use of different methodologies, spatio-temporal scales, different cereals and regions makes a comparison between studies difficult, our findings are consistent with previous works. Using bivariate return periods of combined climate conditions, Zscheischler et al. (2017) have shown how linear models based directly on precipitation and temperature (and not the respective bivariate return period) may underestimate the explained variability of crop yields and that in several countries maize yields decrease with dry and hot conditions. Based on a meta-Gaussian model at the national level, Feng et al. (2019) have also shown that compound dry and hot extremes lead to larger impacts on maize yields than the individual hazards over five major maize-producing countries.

In terms of the relative contributions of drought and heat conditions, a variety of studies at the national scale have found that the response varies from country to country. Feng et al. (2019) have found that China, France and Romania expect higher chances of maize-loss under dry conditions with normal temperatures (rather than under hot conditions with normal

precipitation), while USA and Argentina expect higher chances of maize-loss under hot conditions with normal precipitation (rather than under dry conditions with normal temperatures). In contrast, Zscheischler et al. (2017) have found that countries such as Lithuania, Luxembourg, and the UK, maize yields increase under hot and wet conditions, likely because of the importance of summer precipitation for the crop vegetative cycle and the relatively cooler climate in those countries.

Although previous studies have discussed that maximum temperature might be the best predictor variable for yield variability in most countries (Zscheischler et al., 2017), our study highlights that in Spain crop-loss of wheat and barley is more sensitive to dryness than to hot conditions. This finding agrees with the rainfed practices adopted in the wheat and barley cultivation in Spain. In fact, the nesting structure of the trivariate models adopted in the present study privileges the stronger dependency between yields and precipitation, rather than between yields and temperature or between precipitation and temperature (Figure 5.3). Though irrigated crops typically produce higher yields, the pressure in water resources is already increasing the deficit between water supplies and water demand in Spain (Rodríguez Díaz et al., 2007). Hence, understanding climate risks for rainfed crops is crucial to address the current water management challenges for agricultural practices in Mediterranean regions.

Higher probabilities of crop-loss under drought and/or heat stress are generally expected in the southern region of Spain, in comparison to the northern region (Figure 5.6 and Figure 5.7), in agreement with the higher temperatures and lower rainfall amounts observed in the southern region (IM and AEMET, 2011; Ribeiro et al., 2019a). In the case of wheat-losses, this finding is in agreement with previous work which focused on drought risks for the same crops and the same region (assessed based on remote sensing and hydro-meteorological drought indicators, (Ribeiro et al., 2019c). However, Ribeiro et al. (2019c) identified a higher likelihood of barley-loss with drought in the northern cluster. This discrepancy underlines the importance of addressing the interaction between compound dry and hot conditions and the associated impacts on vegetation. For instance, compound dry and hot conditions have a larger impact on the carbon uptake potential than the sum of the individual impacts (Zscheischler et al., 2014), highlighting the relevance of interactions between multiple stressors.

We found that for barley in cluster 2, drought is the least dominant driver in comparison to the other cereals and regions. Barley in cluster 2 shows the highest difference between drought and

compound dry and hot conditions, and the lowest difference between heat stress and compound conditions (Figure 5.8). This suggests that among both cereals and both regions, barley in cluster 2 is the case where the compound and possibly interacting effects of drought and heat are most relevant. Note that in this case also the CDF's between the dry and hot and dry or hot conditions are more differentiated from each other for the severe and extreme stress (Figure S5.2). This is consistent with a recent study at the province level, which recommended that crop production in Spain should focus more on wheat production given that most provinces displayed lower levels of wheat-loss with drought in comparison to barley-loss (Ribeiro et al., 2019a). This finding is also consistent with Figure 5.6 and Figure 5.7.

The uncertainties associated to the parametric statistical model were assessed with a large number of sampled distributions with the same sample size as the observations. In some of these distributions, drought or heat alone may cause more damage than concurrent drought and heat (lower uncertainty bound is below 0 in Figure 5.8 and FigureS5.3). This highlights the challenges of estimating the likelihood of rare events in two- or three-dimensional probability distribution with limited sample size (Serinaldi, 2013, 2016). For the same reason, the wheat loss in cluster 2 when spring precipitation is below the 5^{th} percentile in Figure 5.7 slightly decreases when the threshold of spring maximum temperature change from the 10^{th} percentile to the 5^{th} percentile (while an increase would be expected like in the other cases). Note that the uncertainties increase with the increasing severity of the compound dry and hot conditions (Figure 5.6 and Figure 5.8) due the rapid decrease of available samples in the corners of the three-dimensional probability distribution. Moreover, following the work by Okhrin and Ristig (2014), here we considered nesting copulas of the same family only, as more complex structures would be difficult to implement in general. Vine copulas might offer an alternative that is also appropriate for higher dimensions (Bevacqua et al., 2017), when considering for instance more driver variables. Nevertheless, in comparison with previous studies based on bivariate models only, statistical modelling based on NAC is a good compromise between complexity and the trivariate dimension.

5.5 Conclusions

The present study assessed how compound drought and heat enhance losses of wheat and barley in two major dryland areas in Spain. Nested Archimedean copulas can successfully model the

trivariate joint distribution between spring maximum temperature, spring precipitation and yields to estimate conditional probabilities of crop-loss under different severity levels of hot and dry conditions. The strongest dependence exists between spring precipitation and yields and is best captured by a Frank copula. The results demonstrate that the probability of crop-loss increases with the severity of compound dry and hot conditions. Furthermore, the likelihood of wheat- and barley-loss increases when drought or heat, respectively, aggravate to compound dry and hot conditions in both regions. Overall, the likelihood of crop-loss in the southern region is larger, in particularly for barley. For both cereals and regions, the likelihood of crop-loss increases more with increasing drought stress than with heat stress, suggesting that drought plays a dominant role in the compound event. The results illustrate the additional value of using trivariate copula modelling to estimate the compounding effects of dry and hot extremes on the risk of crop failure. In operational practice, this research will allow contributing to design supporting tools and provide guidance in the decision-making process in agricultural practices to minimize crop-losses related to climate hazards.

Supplementary Material

Table S5.1 – As in Table 5.2 respecting the possible bivariate pairs of crop yield (u1), precipitation (u2) and maximum temperature (u3), and corresponding Kendall's' correlation (τ). Maximum value of τ denoted in bold and underlined for each cereal and cluster indicating the pair of variables with strongest relationship.

		Cluster 1						Cluster 2					
		τ		G	C	F	J	τ		G	C	F	J
Wheat	$\boldsymbol{C_\theta(u_1,u_2)}$	***0.44***	θ	1.59	0.93	4.75	1.73	***0.51***	θ	1.88	1.37	5.98	2.11
			AIC	-51.43	-47.28	***-69.71***	-35.58		AIC	-71.04	-64.6	***-81.22***	-53.26
			Sn	0.06	0.14	***0.01***	0.17		Sn	0.04	0.11	***0.02***	0.13
	$C_\theta(u_1,u_3)$	0.30	θ	1.28	0.71	2.73	1.27	0.30	θ	1.31	0.53	2.88	1.38
			AIC	-14.3	**-31.71**	-28.51	-4.07		AIC	-13.83	-13.07	**-23.77**	-8.08
			Sn	0.09	0.04	**0.03**	0.18		Sn	0.08	0.1	**0.03**	0.13
	$C_\theta(u_2,u_3)$	0.32	θ	1.4	0.58	3.27	1.51	0.41	θ	1.66	0.77	4.28	1.98
			AIC	-28.45	-21.74	**-38.13**	-20.41		AIC	**-52.05**	-27.27	-48.85	-47.27
			Sn	0.07	0.11	**0.03**	0.13		Sn	0.04	0.14	**0.03**	0.08
Barley	$\boldsymbol{C_\theta(u_1,u_2)}$	***0.44***	θ	1.7	0.92	4.89	1.92	***0.51***	θ	1.94	1.41	6.02	2.21
			AIC	-66.25	-47.07	***-72.18***	-53.18		AIC	-78.79	-68.34	***-81.99***	-61.18
			Sn	0.02	0.13	***0.02***	0.08		Sn	0.03	0.1	***0.02***	0.1
	$C_\theta(u_1,u_3)$	0.30	θ	1.3	0.69	2.77	1.31	0.38	θ	1.46	0.69	3.73	1.61
			AIC	-16.34	**-30.27**	-29.9	-6.11		AIC	-29.33	-22.43	**-38.56**	-21.43
			Sn	0.08	0.06	**0.04**	0.16		Sn	0.09	0.15	**0.04**	0.16
	$C_\theta(u_2,u_3)$	0.32	θ	1.4	0.58	3.27	1.51	0.41	θ	1.66	0.77	4.28	1.98
			AIC	-28.45	-21.74	**-38.13**	-20.41		AIC	-52.05	-27.27	**-48.85**	-47.27
			Sn	0.07	0.11	**0.03**	0.13		Sn	0.04	0.14	**0.03**	0.08

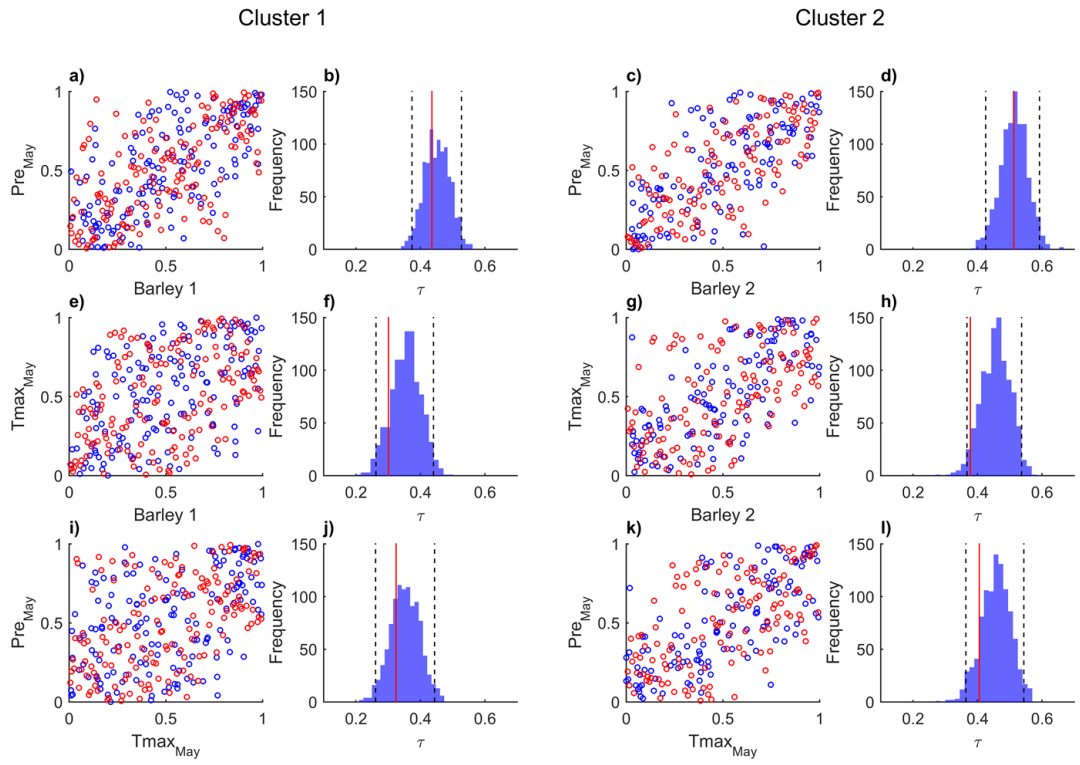


Figure S5.1 - Same as Figure 5.5 for barley.

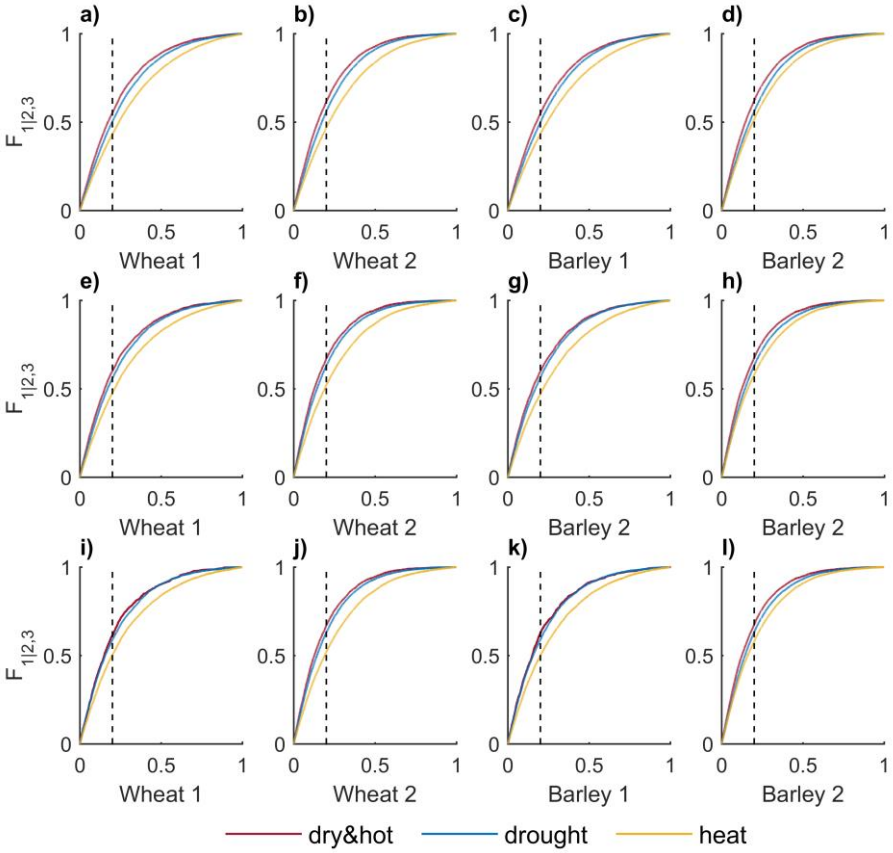


Figure S5.2 - Conditional probability distributions of crop yield anomalies over each cluster under hot (yellow), dry (blue) or compound hot&dry (purple) under moderate (a) - d)), severe (e) - h)) and extreme conditions (i) - l)).

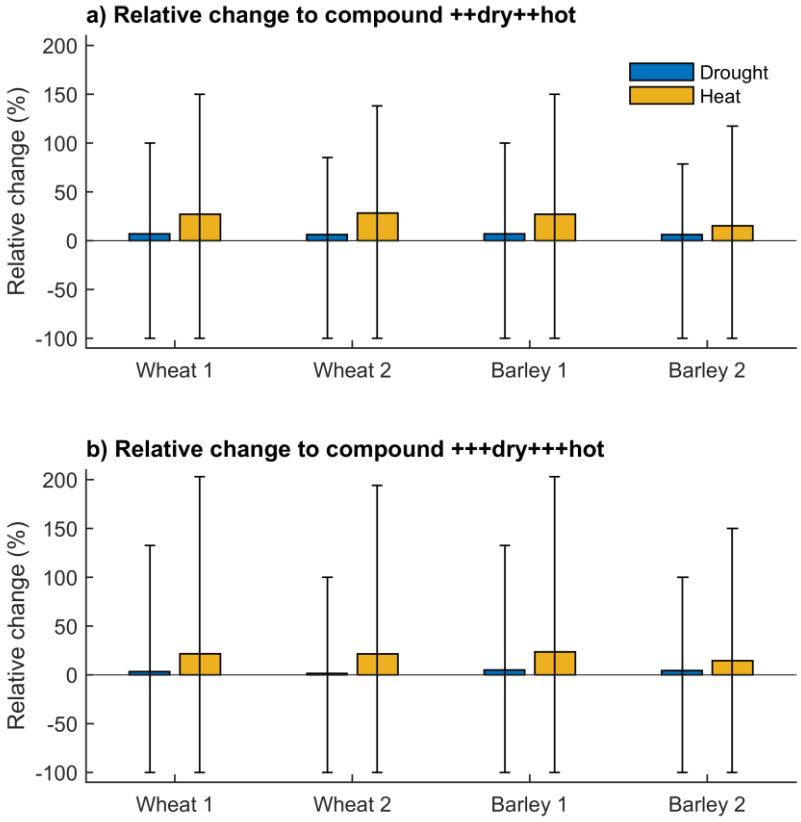


Figure S5.3 - Same as Figure 5.8 for severe (a) and extreme (b) conditions.

Chapter 6

Final remarks

6. Final remarks

6.1 Outlook of the main results

The rationale behind this thesis was centred in an impact-inspired concept of drought-related risk in agriculture in the Iberian Peninsula (IP). This rationale was implemented through the design of suitable statistical models with the aim of establishing a framework for the assessment of drought-related crop-losses. The adopted probabilistic methodologies are expected to effectively contribute to design management options and guide the decision-making process in agricultural practices under drought circumstances.

The outcomes of this thesis were aggregated in four main chapters that aimed to characterize the multivariate, multiscalar, extreme and regional character of drought-related risks in agriculture. To give an outlook of the overall thesis, the following paragraphs address the fundamental findings for every article in chapters 2, 3, 4 and 5, and answer the target questions stated in Chapter 1. Moreover, a brief summary of the main findings achieved in the annexed research work (please see Appendix A) is also addressed in this concluding chapter.

Chapter 2 - Modelling drought-related yield losses in Iberia using remote sensing and multiscalar indices

In Chapter 2, the first question (RQ1) was addressed based on the identification of the most appropriate drought indicators. The characterization of the drought hazard was performed based on two types of indices, namely the hydro-meteorological drought index SPEI and the satellite-based indices VCI, TCI and VHI. Additionally, the relevant timescales of the drought episodes and the moments of the vegetative cycle with highest crop's vulnerability to water stress were also evaluated. The data availability in terms of the rainfed winter cereal yields produced in the IP limited the study, in terms of the crops investigated, to wheat and barley. Afterwards, it was necessary to decide upon the strategy of analysis regarding the risk which each crop constitutes at provincial scale. Therefore, and to answer to the proposed questions, two clusters of provinces dominated by rainfed agricultural practices were identified and selected for further analysis at the regional level. Strong correlations between the crop yield and the referred indices were found during crop intermediate growth stages (spring and summer), rather than in the initial growth stages (autumn/winter).

Based on correlation analysis, the results suggested that yield loss is associated with moisture stress (low values of VCI) during early-spring and with heat stress (low values of TCI) during early-summer. The dominant time-scales at which drought influences the crop yield corresponded to longer time-scales (6 to 12 months) throughout January to June, and a pronounced impact is verified during the spring months (April to June) at the short time-scales (1 to 6 months).

Other question of interest (RQ2) was related with the skill of the several drought indicators in predicting yields before harvest and events of crop-loss. To this purpose, the time scales and months of SPEI, together with the weeks of VCI, TCI and VHI, better related with crop yields were chosen based on stepwise regression to establish further MLR and ANN models using cross-validation to estimate the yield response to drought conditions.

The selection of significant predictors indicated that the southern sector crop yield is better inferred from the drought index SPEI information, while the remote sensing of vegetation is more relevant in the northern sector crop yield. The results of both the statistical approaches suggested the potential of modelling drought-related yield losses, and minor improvements were found by the ANN in comparison with the MLR methods.

Chapter 3 - Probabilistic modelling of the dependence between rainfed crops and drought hazard

While the rationale behind the previous chapter lied on the effects of explanatory variables (drought indicators) on a response variable (crop yield anomalies), this chapter moved forward towards the designing of a comprehensively framework for a risk analysis of drought-related crop-loss, and therefore trying to answer to RQ3 and RQ4. Given the suitability of using the formerly identified two clusters of provinces for the analysis of drought-related impacts in agriculture at the regional level, in this chapter the same area was considered for analysis. For the same reason, and for sequential purposes, the crop and drought hazard data used in Chapter 2 have been incorporated in this study for the probabilistic modelling of the dependence structure between droughts and crops based on copula functions with different types of tail symmetries and asymmetries.

The core idea to be explored within this chapter was essentially methodological, in order to design a framework that could serve as technical basis for the future steps based on the statistical

inference of the drought-related risks in agriculture with an intuitive physical interpretation. The methodology is based on a bivariate copula approach to model the joint probability distributions between drought conditions and crop yield anomalies and derive drought-related conditional probabilities of crop-loss. The suitability of this approach was verified by comparing conditional with unconditional probabilities of crop-loss, i.e., the probability of an event of crop-loss without conditioning to the occurrence of a drought event with the probability of an event of crop-loss conditioned to the occurrence of a drought event. In addition, in terms of conditional probabilities, the probability of drought-related crop-loss was compared with wetness-related crop-loss.

The bivariate copula models suggested that the dependence structure between crop yield anomalies and drought conditions is mainly asymmetrical, which suggested the presence of dependence among extreme values of yield and drought indicators. Afterwards, the derived risks of agricultural impacts suggested that the use of unconditional probabilities of crop-loss may lead to the underestimation of the agricultural risks.

Moreover, the conditional probabilities of crop-loss indicated that the risk of wheat-loss and barley-loss increases when drought events aggravate from normal/wet to moderate/severe conditions. From a decision-making point of view, the estimations of drought-related crop-loss indicated that the risk of barley-loss is more likely to occur in the northern sector, while the risk of drought-related wheat-loss is more likely in the southern sector, advising that sowing in cluster 1 (cluster 2) may possibly be more focused on wheat (barley).

Chapter 4 - Copula-based agricultural drought risk of rainfed cropping systems

In the previous chapter, the use of bivariate copulas for the analysis of the co-occurrence of dry and low-yield events have shown the added value of applying this type of methodology for the estimation of drought-related crop failure. A further step was taken, and the same statistical design was applied to map the drought-related risk in agriculture at the province level of the IP, providing a clear and interpretable measure of risk for decision support. Additionally, the analysis was extended for a longer period, allowing also to distinguish between different classes of drought severity (RQ5).

Results stressed that in most of the provinces, the dependence between drought and yield extremes is mainly in the lower values, when drought is characterized by VCI and SPEI. In

contrast, the joint behavior between the TCI and wheat suggested greater probabilities of joint high extremes.

In general, the drought-related crop-loss increased with drought severity in most of the provinces, with barley exhibiting greater agricultural drought risk in comparison to wheat. Nevertheless, minor wheat- and barley-losses can still be expected during normal or wet conditions, stressing the complexity of the interactions between the agricultural systems and the climate.

Chapter 5 - Risk of crop-failure due to compound dry and hot extremes estimated with nested copulas

The results of previous chapters consistently pointed to the fact that wheat and barley are crops vulnerable to drought conditions, and that barley was the cereal at greatest risk of induced drought damages. This suggested that crop production in the IP would beneficiate of lower crop-losses if it could be more focused on wheat given the lower levels of crop yield loss attributable to droughts using the applied methodology. Nonetheless, some losses of both cereals were still expected under non-drought conditions, and in this context, a question of interest (RQ6) addressed in Chapter 5 was whether hot extremes amplify the impact of droughts on crops if they occur at the same time. Ultimately, in line with a compound event perspective of the climate-related agricultural risks, this part of the PhD research intended to understand the role played by drought in the occurrence of compound events in the agricultural systems over the IP.

While previous chapters analysed the bivariate dependence between the drought hazard and crop yields, this chapter analysed the trivariate dependence structure between maximum temperature, precipitation and crop yields to study the impacts of compound dry and hot extremes on crop yields. The resulting statistical model allowed the sampling from the full trivariate distribution and to estimate the increase in likelihood of crop failure during concurrent hot and dry conditions compared to hot or dry conditions individually.

The results suggested that the probability of crop-loss increases with the severity of the compound hot and dry conditions, and that drought plays the dominant role in crop-loss due to the compound event. In consensus with previous chapters, the likelihood of barley-loss is

slightly higher than wheat-loss, and the southern cluster of provinces dominated by rainfed agriculture also suggested more susceptibility to compound dry and hot extremes in agriculture.

Appendix A - Drought-related hot summers: a joint probability analysis in the Iberian Peninsula

The steps carried out during these four years allowed for the accomplishment of the PhD work plan and also for the development of complementing material. In this way, an appendix section addresses some parallel work to this thesis (please see Appendix A). The fundamental questions addressed here is how summer hot extremes are exacerbated by the occurrence of droughts in the IP. Therefore, it was proposed to extend the knowledge on this combined effect to the occurrence of drought-related hot summers in the IP in Appendix A, considering the previous conclusions that the analysed impacts of droughts increase with the interaction with extreme temperatures (Chapter 5). The results showed that the transition from previous wet to dry regimes increases substantially the probability of exceeding summer extreme temperatures, which could contribute to mitigate the associated magnified impacts. This is in accordance with previous works which show a strong association and a certain predictability between preceding drought conditions and summer extreme temperatures in the IP. However, previous studies lack to address the probabilistic risk of exceeding extreme temperatures under drought conditions, which were addressed in Appendix A. Through the assessment of the conditional probabilities of drought-related hot summers a further step was taken, adopting a copula-based methodology, which was a cross-sectional concept in this thesis explored in chapters 3, 4 and 5.

6.2 General discussion, limitations and future work

The main core of this thesis was to better understand, assess and quantify the risk of drought impacts in agricultural system. In this sense, a comprehensive framework was proposed, relying on the use of classical drought indicators together with drought related vegetation indices obtained from satellite data. Additional data was required namely land cover classification and crop yields. Innovative techniques were tested to analyze dependence of extreme values and conditional dependencies, namely the copula theory.

An initial effort was made to identify the direct impacts of the drought phenomena on agricultural systems, here defined in terms of reduced crop yields. Hence, the depiction of the

cause (drought) and consequence (crop-loss) is fundamental for risk analysis. For this reason, the first steps of the conducted research (Chapter 2) focused on the response of rainfed crop yields to drought and in the predictability of crop yield losses given the previous state of the drought conditions before the harvest. The main findings of the research conducted in Chapter 2, such that the greatest influence of short timescale droughts on crop yields is observed during the spring and early-summer, are supported by several previous works (Vicente-Serrano et al., 2006; Páscoa et al., 2017b; Peña-Gallardo et al., 2019). In particular, (Páscoa et al., 2017b) shows that May is the month with largest correlations between SPEI and wheat yields at the province scale (similarly to the findings in Chapter 5), and results from (Peña-Gallardo et al., 2019) are similar with the added benefit of using a district scale allowing a finer spatial resolution to characterize the response of crop yields to drought indices. In comparison to these previous studies, the research presented in this thesis took a step forward by focusing on crop-loss events more than in representing the interannual yield variability. Although Páscoa et al. (2017b) had analysed the simultaneous occurrence of years of low yield anomalies and dry events, it lacked the modelling of these concurrent extreme events and a probabilistic risk analysis.

While primarily it was addressed in Chapter 2 the capability of predicting the events of crop-loss based on a pool of statistical significant predictors (drought indicators), the next steps addressed the probability of occurring a crop-loss event given the occurrence of a dry event. The goal intended to attribute a probabilistic risk measure to drought impacts in rainfed agriculture. In contrast to deriving an empirical equation, the rationale was to estimate the joint probability distribution (copula) of the response and explanatory variable(s) and derive conditional probabilities instead. One of the advantages of using the designed framework relates with the simplification of the risk terminology in respect to the vulnerability and exposure analysis, and quantifying the risk of yield reduction induced by droughts in a comprehensive way. For instance, in comparison with the approach by Murthy et al. (2015) and Alonso et al. (2019), which are crop-generic and require a full list of weighted indicators to identify the agricultural areas most vulnerable to drought, the simple use of conditional probabilities may be more intuitive for the end user. The referred works propose including a higher number of variables using composite index methods, however being based on static data over time. Conversely, the methodology used in this thesis uses time series data, which is an advantage.

Moreover, analysing both the probability of the hazard (drought) and the consequence (crop-specific yield loss) over the major rainfed agricultural areas (Chapters 2, 3 and 5) provides data-based and crop-specific tools to decision-makers which allow for guided operational decisions. In particular, and with the increase of crop yields in sight, results highlighted that sowing wheat is advantageous in preference to barley.

On one hand, this thesis results suggest that wheat may be a more drought-tolerant crop, in comparison to barley. On the other hand, barley is usually broadly acknowledged to be a more drought-tolerant crop in comparison to wheat worldwide. However, it is still a matter of discussion, particularly in the Mediterranean environments, with very different results being published (Austin et al., 1998; Cossani et al., 2009; Carvalho et al., 2014). For instance, an experimental study by Cossani et al. (2009) in the IP area of the Ebro Valley, did not identify a clear overperformance of barley over wheat under lowest-yielding conditions. Another experimental study in Mediterranean cultivars, found that drought decreased grain yield more for barley (47%) than durum wheat (30%) (Carvalho et al., 2014). However, a site study by Austin et al. (1998) in the Ebro Valley using correlation analysis supports that barley is better adapted than wheat to very dry rainfed conditions, thus having lower risk of crop-loss. Similarly, Peña-Gallardo et al. (2019) also found differences in the magnitude of the correlations between the drought indices and these cereals over the IP, with wheat yields showing greater correlations than barley yields, suggesting that barley is less dependent on water availability. It is worth mentioning that the studies over the Ebro Valley are specific to that region, which is a semi-arid region with dominant drought and heat stress, in comparison to other regions of the IP as shown in the Appendix A of this thesis and in previous studies by several authors (Vicente-Serrano and Cuadrat-Prats, 2007; Vicente-Serrano et al., 2014; Ojeda et al., 2016).

In this respect, a recent turning point in risk analysis relates with the assessment of extreme impacts as the consequence of the interaction among multiple drivers and/or hazards (compound events). The use of the proposed compound framework to characterize the additional impacts of dry and hot extremes on crop yields identified drought as the main driver of crop-loss, even when interacting with extreme hot conditions. In addition, the results show that barley in the southern region is more susceptible to the interacting effects between drought and heat. This is expected to be exacerbated as, based on a recent global study which evaluates

the effects of concurrent drought and heat extremes under climate change scenarios, increasing temperatures may cause substantial decreases in barley yields worldwide (Xie et al., 2018).

The achieved outcomes and the identified research gaps suggested several new possible research lines which are briefly discussed below:

1) One of the most evident follow-ups is studying future risks of rainfed yield shortfall under climate change scenarios. Currently, a study is under preparation with the collaboration of several IDL researchers, focusing mainly on the drought hazard and using data from different General and Regional Circulation Models (GCM's or RCM's). The fundamental question to be addressed is how drought conditions will evolve in IP for both 1.5º and 2ºC warming targets for two future climate scenarios (RCP4.5 and RCP 8.5). In this respect, a question of interest is if future crop failure is expected to increase with the additional 0.5º of warming from 1.5º to 2ºC.

2) The study in Appendix A, which focus of drought-related hot summers in the IP, constitutes a natural progression of the current line of research. In particular, the focus on the multi-hazard analysis and the resulting multi-impacts, from an historical and future perspective which entail important features for risk assessment and disaster reduction frameworks. Such multidisciplinary perspective involves the development of new hazard indicators, identification of their drivers, identification of the trigger conditions causing the extreme impacts and collection of ecological impact database. For instance, the integrated analysis of the cascading events in the agricultural and forest systems has caught the attention of several scientific projects facing the challenging of linking multiple weather drivers to multiple extreme impacts. In this regard, further improvements to the present study might also be including other factors that could influence the climate–yield relationship, such as minimum temperatures, cold spells, soil moisture and wind speed. From a remote sensing perspective, additional variables, such as land surface temperature and evapotranspiration together with biophysical parameters such as the leaf area index, the vegetation biomass or the net primary production could be included. In addition, pest occurrence and use of pesticides are directly influenced by climate conditions and are also expected to continue to be affected with climate change (Delcour et al., 2015; Lechenet et al., 2016). Taking into account other factors rather than crop's yield such as soil composition, pests, diseases would

guide farmers and help them to adopt more sustainable practices. In terms of methodological settings this would require higher dimensional models and a larger sample size.

3) From another point of view, the association between events of lowest yield and the prevailing weather patterns or large atmospheric circulation modes in the North Atlantic region are possible lines of complementary research to identify the large scale drivers of crop-loss. Low-frequency modes can be interpreted as sources of predictability because the ability to predict well the low-frequency modes, may lead to a good prediction of regional features, such as droughts, and hence may lead to a good prediction of the consequent impacts, such as crop-loss. Previous studies have already shown significant relationships with wheat yield in the IP and atmospheric patterns, such as the North Atlantic Oscillation (NAO) (Gouveia et al., 2008; Capa-Morocho et al., 2016b). Other large-scale drivers of climate variability such as Sea Surface Temperature (SST) and other teleconnections such as the El Niño have also shown to be associated to the winter cropping systems in the IP (Capa-Morocho et al., 2016b). At the global scale, a study by Anderson et al. (2019) identified the El Niño as a major influencing factor of global crops and responsible for the largest recorded synchronous maize failure in 1983. In addition to these studies, the multivariate statistical downscaling joining regional-scale events driven by larger-scale systems adopting a compound event perspective could bring significant contributions to the modelling of crop yields and prevention of major crop failures.

4) In line with economic perspectives in a changing climate, climate-related synchronous crop failures across multiple regions of the world could also help to balance global trade networks to support food security. Climate-related crop failures concurrent across multiple regions of the world have already been addressed in recent studies (Anderson et al., 2019; Gaupp et al., 2019) as if poor harvest occur at several places simultaneously the global food system may be heavily compromised. Based on spatial dependence structures, Gaupp et al. (2019) shows that wheat-losses in EU may be exacerbated by increasing temperature in Australia in the case of a drought during the wheat-growing season, due to the trade networks. In contrast, maize-losses in EU induced by hot conditions may be mitigated by a precipitation regime favorable to maize yields in

Brazil (Gaupp et al., 2019). The consideration of the relationships associated with exportations and importations of wheat and barley in the IP could entail potential goals for future work due to the major importance of cereals as staple food for the populations and for securing global food, particularly in a changing climate.

5) In fact, the definition adopted in this thesis to outline drought-related risk in agriculture neglected a socio-economic analysis to guideline stakeholders and policymakers. Relevant factors regarding the spatial distribution of population and consequent food demand and domestic prices, cereal's market prices (price for sellers rather than consumers), farmers economic income, machinery and other operational costs, national Gross Domestic Product (GDP), among others, would require the integrated use of several datasets with different temporal and spatial scales. A number of authors have analysed the socio-economic impact of droughts in the Mediterranean and IP, namely the pressures caused by the water demand (Iglesias et al., 2011), the livestock systems (Iglesias et al., 2016) and irrigated agriculture (Gil et al., 2011). A wide variety of methodologies for assessing the drought impacts in the socio-economic sector have been considered, such as econometric (Gil et al., 2011) and bio-economic models (Gbegbelegbe et al., 2014) towards the contribution for the financial, insurance and reinsurance markets.

6) Finally, scientific efforts are required towards the progress of food security in developing countries, namely in the African continent, in view of the development of global climate-smart agricultural practices. In addition to the cereals discussed in this thesis, other major crops are of great importance over these countries, for both self-sufficiency and economy, such as cotton, coffee, sugar, soybean, among others. Hence the application of the methodology proposed in this thesis over other regions and crops, entails promising future lines for research and cross-cultural scientific exchange.

6.3 Deliverables during the thesis

Through this doctoral research period, three first-author peer-reviewed articles were published in the context of the present PhD (chapters 2, 3 and 4), and two first-author articles (chapter 5 and Appendix A) are under revision during the writing of this thesis. The results of this research

were presented at different national and international conferences as oral and poster communications (see Curriculum Vitae). Although not initially planned, a 3-months Short-Term Scientific Mission (STSM) was done in the Climate and Environmental Physics research group in the University of Bern, funded by the COST Action CA17109 which contributed to the elaboration of a research article (Chapter 5). In addition, collaborations with several IDL researchers and with the network of the CA17109 have resulted in the participation of parallel research not related to the PhD itself but related to the topic (see Curriculum Vitae). Particularly, a study by Péres et al. (2020) about the links between air temperature and mortality in Brazil and a paper under preparation by Russo et al. (to be submitted in 2020) about the climate change impacts in drought conditions in the IP. Competitive grants such as the Roland Schlich Travel Support and the IDL Travel Award for EGU 2020, and three grants from the COST Action CA17109 (for the Workshop on Correlated Extremes, Training School on Statistical Modelling of Compound Events and STSM) were obtained.

6.4 Take-home message

In summary, the results of this thesis pointed to the fact that wheat and barley are crops vulnerable to drought conditions and, in general, barley presents a higher drought-related risk compared to wheat. The results showed the good performance of drought indicators in predicting crop-losses and that the drought-related risk in agriculture increases with the severity of drought conditions. Moreover, the risks related to drought increase with the interaction with extreme temperatures. Nevertheless, drought is the dominant factor of wheat- and barley-losses induced by simultaneous drought and heat conditions.

In order to contribute to agricultural systems more resilient to extreme weather events, this thesis has grasped the importance of addressing drought-related risks adopting an impact-centric perspective and taking the most of available resources for drought monitoring and suitable statistical methods.

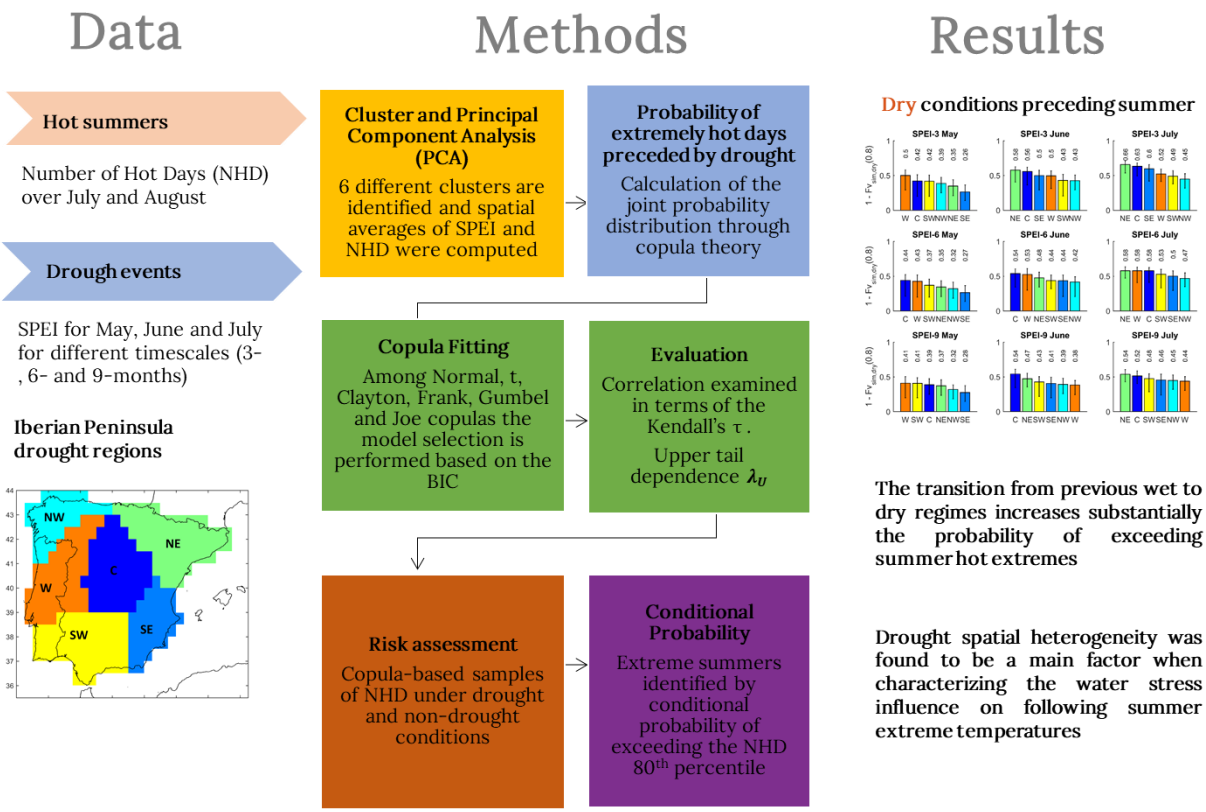
Appendix A

Drought-related hot summers: a joint probability analysis in the Iberian Peninsula

Ribeiro, A.F.S., *Russo, A., Gouveia, C., Pires, C A. L., Drought-related hot summers: a joint probability analysis in the Iberian Peninsula, Submitted to Weather and Climate Extremes*

Andreia F. S. Ribeiro contributions*: statistical analysis and interpretation of the results (principal component and cluster analysis, correlation analysis, copula-based analysis); production of the figures, manuscript writing and coordination.*

Graphical abstract



Highlights

- The dependence between drought and hot events is identified for most of the Iberian Peninsula's regions
- The transition from previously wet to dry regimes increases substantially the probability of exceeding the number of extremely hot days, reinforcing that summer hot extremes may be induced by previous soil moisture deficits
- Results are not spatially homogeneous over the study region regarding the influence of water deficits on following summer extreme temperatures
- North-eastern, western and central regions were found to be the regions more prone to summer hot extremes induced by dryness

Abstract

Droughts and hot extremes are major sources of risk to several socio-economic activities and their impacts are expected to increase under future global warming. Moreover, the simultaneous or sequential occurrence (compound events) of different climate extremes may lead to the amplification of the associated impacts. Even though the latest efforts in assessing hot and dry extremes and their interactions, the development of models describing the joint behaviour of climate extremes is still a challenge. To contribute to the understanding of these compound events, the probability of summer extremely hot days in the Iberian Peninsula (IP) being preceded by drought events in spring and early summer are assessed based on their joint probability distribution through copula theory. Drought and extremely hot events were identified respectively using the Standardized Precipitation Evaporation Index (SPEI) in the months of May, June and July for different timescales (3-, 6- and 9-months) and the Number of Hot Days per month (NHD) summed over the months of July and August. The dependence structure between SPEI and NHD was very well identified for the most of the IP's regions by asymmetrical copulas with upper tail dependence (except in north-western regions), suggesting that compound hot and dry extremes are strongly associated. The results show that the transition from previous wet to dry regimes increases substantially the probability of exceeding summer extreme hot days, depending on the region and the drought timescale and target month. The results suggest a spatial heterogeneity over the IP when characterizing the influence of water deficits on following summer extreme temperatures, whereas north-eastern, western and central regions were found to be the regions more prone to summer hot extremes induced by dryness, in contrast to southwestern, north-western and south-eastern regions. This approach can ultimately be a valuable tool for responsible authorities to mitigate the impacts magnified by the interactions between the different hazards.

Keywords: Droughts, Hot extremes, Compound events, Standardized Precipitation Evaporation Index (SPEI), Number of Hot Days (NHD), Copulas, Conditional probability

A. Drought-related hot summers: a joint probability analysis in the Iberian Peninsula

A.1 Introduction

Drought and heat-related extremes (e.g. heatwaves, warm spells, hot days) are among the most influential climate hazards (Yuan et al., 2016; Zampieri et al., 2017; Lu et al., 2018) as they lead to a variety of impacts, such as exacerbation of fire risk and crop damage, causing several economic losses and adverse effects in human health and mortality (Gouveia et al., 2016b; Mazdiyasni et al., 2017; Russo et al., 2017; Zampieri et al., 2017). The last IPCC reports on extreme events point out that unprecedented risk to humans and ecosystems are expected in a changing climate due to changes in precipitation and temperature regimes and extremes (IPCC, 2012, 2018), particularly if rapid and far-reaching transitions are not met in several economic sectors and areas (e.g. land, energy, industry, buildings, transport and cities) (IPCC, 2018).

Several studies have stressed the role played by the interplay between multiple climatic extremes, which may exacerbate the impacts of individual hazards (Zscheischler and Seneviratne, 2017; Zscheischler et al., 2018). For example, in the US although there is no significant trend in drought from 1960 to 2010, a substantial increase of concurrent droughts and heatwaves is observed during that period (Mazdiyasni and AghaKouchak, 2015). In Europe, the 2003 and 2010 extreme heatwaves were concurrent with serious drought conditions, causing more damage than extreme temperatures or extreme dryness would have caused individually (Hauser et al., 2015). In southern Greece, the extreme fire season of 2007 was driven by two complementary climatic extreme drivers: a major drought in preceding months and two major heatwaves in July and August (Gouveia et al. 2016). Also in the 2010 European summer, the previous conditions of low soil moisture in spring strongly amplified the magnitude of the devastating heatwave (Hauser et al., 2015). Similarly, in the case of the European heatwave in 2003, if soil moisture in previous spring levels had not been as close to climatology as they were, the temperatures could have been much higher (Whan et al., 2015). More recently, the catastrophic fire seasons of 2019 in Greece (NASA Earth Observatory, 2019a) and 2019/2020 in Australia are also pointed out to be associated with a drought exacerbation of summer hot conditions (NASA Earth Observatory, 2019b).

The processes involved in the soil moisture–temperature coupling and feedback have been addressed in a growing number of works (Zampieri et al., 2009; Seneviratne et al., 2010; Mueller and Seneviratne, 2012; Miralles et al., 2014, 2019; Vogel et al., 2018), with a clear relationship between dry/wet extremes and the frequency of hot extremes being identified in Southeastern Europe (Hirschi et al., 2011; Russo et al., 2019), in Iberian Peninsula (IP) (Mueller and Seneviratne, 2012; Russo et al., 2019), at almost all South America, Indonesia and Malaysia and extensive areas in North America (Mueller and Seneviratne, 2012). Consistent signal was identified, both with soil-moisture and soil-moisture proxies (Standardized Precipitation Evapotranspiration Index, SPEI, and Standardized Precipitation Index, SPI), with summer heat predictability being conditioned by preceding dry/wet anomalies (Russo et al., 2019), summer circulation anomalies (Quesada et al., 2012) and advected sensible heat which further strengthen local land–atmosphere feedbacks via soil desiccation (Schumacher et al., 2019). In this context, it is becoming vital to make use of a compound event approach for understanding the extreme impacts and investigate opportunities for predictability towards the mitigation of the consequences (Zscheischler et al., 2018).

This aim is even more important under the projected warming and drier last decades of the 21st century (IPCC, 2019), which may enhance the occurrence of more extreme compound events (Lu et al., 2018). In addition, the combined effects of the projected global warming have been particularly pronounced in specific regions, as is the case of the Mediterranean areas (Giorgi and Lionello, 2008). Among the Mediterranean areas, the IP is an outstanding example of how precursor soil-moisture deficits conditions are strongly related to the extreme hot temperatures in the following summer (Russo et al., 2019).

The present work goes a step further, by addressing a key property of compound extremes related to the existence of multivariate dependence structures between the involved variables (Hao et al., 2018b; Feng et al., 2019) through the use of a copula approach. The use of copulas is among the most recently applied techniques in multivariate dependence modelling in climate studies (Zscheischler et al., 2017; Ribeiro et al., 2019a, 2019c), and a couple of recent works have adopted copula-based methods to model the joint behavior of hot and dry extremes (Hao et al., 2017; Feng et al., 2019). Nevertheless, other methods such as the counting of number of simultaneous/consequential occurrences of multiple extremes (Hao et al., 2018c; Wu et al., 2019) and the use of an indicator approach such as the Standardized Dry and Hot Index (SDHI) (Hao et al., 2018b) have also been applied to characterize compound hot and dry extremes. The

present preference for adopting a copula-based approach is explained by the unique characteristics of copulas which allow for a full characterization of the dependence between variables, even with complex relationships, aside from the shape of marginal distributions (Nelsen, 2013; Maity, 2018).

The characterization of joint occurrence of extreme temperature and dryness is here performed based on the number of hot days (NHD) (Fischer et al., 2007; Zhang et al., 2011b) and on the SPEI (Vicente-Serrano et al., 2010), respectively. The SPEI was used aiming to include the effect of evapotranspiration on drought monitoring, which is particularly relevant in the context of global warming (Beguería et al., 2014) and to take advantage of the index's multiscalar character. Moreover, in the IP the occurrence of dryer conditions is generally better characterized by SPEI rather than by simplest indices solely based on rainfall records, given the ability of including the impacts of warming processes (Vicente-Serrano et al., 2011). Nonetheless, the projected warmer climate is also likely to be enhancing the land-atmosphere feedbacks (Seneviratne et al., 2010; Miralles et al., 2019), and the precursor effects of low soil moisture availability also play a role on the following hot extremes. In this way and under the perspective of compound events, namely when feedback mechanisms are involved, the rationale of the present work is to highlight the adoption of statistical methods accounting for the joint behavior between the extreme variables, to contribute for the design of adaption measures. For this reason, the goal of the present work is to quantify the likelihood of occurrence of summer extremely hot days, preceded by the occurrence of droughts, preserving the dependence structure between NHD and SPEI in the IP for the period between 1950 and 2014.

A.2 Materials and methods

A.2.1 Data and study area

The NHD was defined as the number of days with maximum temperature exceeding the 90th daily percentile centred on a 5-day window. For more details please see Zhang et al. (2011b). NHD were computed based on regularly gridded (0.5° resolution) values of daily temperatures for 1950-2014 period from the ECAD-EOBS v14 daily dataset (Haylock et al., 2008). To avoid

time discontinuities, the NHD were summed up at each grid point over the two hottest months in the IP, July and August.

The SPEI was calculated based on precipitation and reference evapotranspiration from the CRU TS 4.01 database (0.5° resolution) for the 1950-2015 period (Harris et al., 2014) for 3 timescales (3-, 6-, 9-months). SPEI computation was performed using a log-logistic distribution for statistical fitting (Vicente-Serrano et al., 2010) and a Normal kernel function to accumulate the previous months. SPEI was used as proxy for surface moisture deficits to assess the impact of these deficits on the occurrence of subsequent hot days in the following summer months. Hence, the SPEI values for the months of May, June and July were considered to include the period preceding the hottest months and overlap the month of July to account for co-occurrent compound events.

Due to the high spatial variability of drought conditions in the IP, here we determine main spatial-temporal drought modes based on SPEI over the IP applying a Principal Component Analysis (PCA) considering the months of May and June and the 3 timescales (3-, 6-, 9-months) (Figure SA.1, Figure SA.2 and Figure SA.3). For each timescale, the four principal components explaining most of the variance were rotated based on the varimax method (Hannachi et al., 2007 and references therein) and the main 3 were retained, leading to 9 different drought modes resulting from the 3 timescales. The 9 drought modes are considered for a k-means cluster analysis (Russo et al., 2015b) in order to capture different homogenous drought regions over the IP. Based on the cluster analysis six different drought clusters are identified and spatial averages of SPEI and NHD were computed over each regional cluster.

To each pair of SPEI and NHD over each region the correlation is examined in terms the Kendall's τ, which is a rank correlation test measuring the level of dependence between the datasets (i.e. independence test based on τ). The results of the Pearson correlation coefficient are shown in the Figure SA.4, assuming linearity and that both variables are normally distributed, while Kendall's τ is ranked based avoiding the assumptions about the marginal distributions, similarly to the copula analysis described below.

A.2.2 Joint probability analysis

Among multivariate analysis approaches, copula functions are becoming quite popular (Mirabbasi et al., 2012; Lee et al., 2013; Li et al., 2015). The Sklar's theorem states that a joint probability distribution can be split into its univariate margins and a copula which describes the dependence between the margins. Mathematically, given two correlated variables, X and Y, with marginal distributions $F_X(x)$ and $F_Y(y)$, uniform on the interval [0,1], a copula function C links these distributions to their joint probability distribution $F_{XY}(x, y)$ as follows:

$$F_{XY}(x, y) = C\big(F_X(x), F_Y(y)\big) \tag{A.1}$$

Here, X denotes the SPEI, Y the NHD and (x, y) are pairs of respective observations. Copula functions show a great flexibility in modelling the dependence between variables with complex relationships and are adequate tools for modelling tailed events (extremes) in multivariate distributions (Mazdiyasni et al., 2017). There is a range of copula families described in the literature allowing the modelling of several different shapes of radial asymmetry or symmetry, and different patterns of tail dependence (Figure A.1).

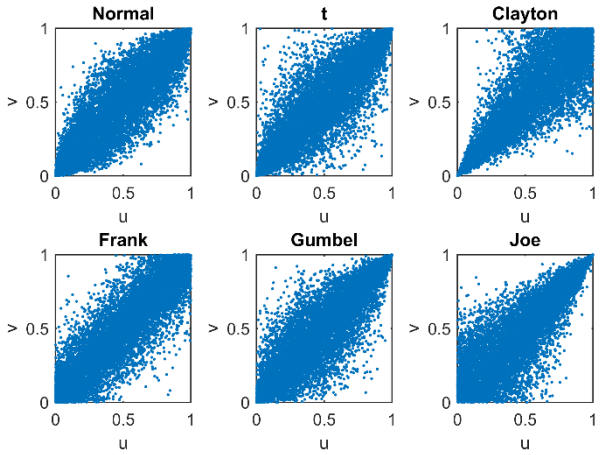


Figure A.1 - 10000 sample points generated from the employed copulas illustrating the dependence structures using as example $\tau = 0.7$.

To each pair of SPEI and NHD we fit six different copulas: Normal, t, Clayton, Frank, Gumbel and Joe copulas (Figure A.1 and Table A.1). These well-documented copula functions belong to two distinct classes of copulas: Elliptical (Normal and t) and Archimedean (Clayton, Frank, Gumbel and Joe). The Archimedean copulas have an explicit formula with only one parameter (Table A.1) and are quite popular given their ability in capturing a wider variety of joint dependence structures. The Clayton, Gumbel and Joe copulas describe an asymmetric tail

behavior whereas Clayton copula can model lower tail dependence and Gumbel and Joe are able to model upper tail dependence (Figure A.1). In contrast, the Normal, t and Frank copulas can capture symmetric dependencies, with the difference that the t copula allows dependence in the extremes in both lower and upper tails (Figure A.1). Note that the Gumbel copula is both an Archimedean and extreme-value copula, and for this reason we consider also the Joe copula to capture upper tail dependencies. In addition, we aim to model the full distribution to capture all the events, and not only the extreme events.

Table A.1 - Equations of the copula functions, where u and v are univariate variables, Φ^{-1} is the inverse of standard CDF, ${t_{df}}^{-1}$ is the inverse t–student CDF, df is the degree of freedom, ρ and θ are dependence parameters.

Family	**Joint Cumulative Distribution Function** $C(u,v)$	**Parameter range**
Normal (N)	$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} exp\left(-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right) dudv$	$-1 \le \rho \le 1$
t	$\int_{-\infty}^{t_{df}^{-1}(u)}\int_{-\infty}^{t_{df}^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left(1 + \frac{u^2+v^2-2\rho uv}{df(1-\rho^2)}\right)^{-\frac{df+2}{2}} dudv$	$-1 \le \rho \le 1$ $df \ge 1$
Clayton (C)	$(u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}}$	$0 \le \theta$
Frank (F)	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$\theta \ne 0$
Gumbel (G)	$e^{[(-lnu)^\theta+(-lnv)^\theta]^{\frac{1}{\theta}}}$	$1 \le \theta$
Joe (J)	$-\left\{(1-u)^{-1+\theta}\left[-1+(1-v)^\theta\right]\left[(1-u)^\theta+(1-v)^\theta - (1-u)^\theta(1-v)^\theta\right]^{-1+\frac{1}{\theta}}\right\}$	$1 \le \theta$

In this work, the copula fits are performed based on a semiparametric method, where the sample data is first transformed into uniform variables (u, v) using a nonparametric estimation (rank based) of the margins, and afterwards the copula parameters (θ) are estimated based on maximum likelihood. Due to the negative character of the correlation between SPEI and NHD (Russo et al., 2019), here we have considered margins of the symmetric (mirrored) SPEI data to simplify the copula modelling. Accordingly, drought conditions correspond to positive values of uniform SPEI and wet conditions correspond to negative values of uniform SPEI (the mirrored SPEI values). The copula model selection is performed based on the Bayesian

information criterion (BIC), which is similar to the Akaike's information criterion (AIC), but penalizes the number of parameters more heavily. In the case of t copula, that have two parameters controlling the tail dependence, AIC and BIC disagrees, as a t-copula would be selected based on AIC but not based on BIC (Table SA.1). The goodness of fit is performed by comparing with the respective empirical copula based on the Cramer-von Mises distance using a parametric bootstrap (Genest and Remillard, 2008).

Once the best fit for each SPEI and NHD combination is performed, uniformly distributed data is sampled from the selected copula models allowing for estimating the conditional probability of exceedance extreme summer NHD values when preceded by drought conditions and compared to when preceded by normal/wet conditions. Based on the estimated joint distributions, 10000 samples of SPEI and NHD denoted u_{sim} and v_{sim}, respectively, are generated preserving the dependence structure between the variables. Note that the samples hold the mirrored values of SPEI, wherefore, drought conditions correspond the upper quantiles and wet conditions correspond to the lower quantiles. Hence, and since the simulations are uniform in the range [0,1], here drought is identified when u_{sim} is equal or above 0.8 (equivalent to the quantile 0.2 of SPEI (-0.84) according to the conventional SPEI severity level associated with moderate drought) and normal/wet conditions when u_{sim} is below 0.8. Afterwards, the correspondent NHD samples under drought conditions ($v_{sim,dry}$) and the NHD samples under normal/wet conditions ($v_{sim,wet}$) are considered. The conditional survival functions $1 - F_{v_{sim,dry}}$ and $1 - F_{v_{sim,wet}}$ are easily obtained by ordering the uniform samples ranging between [0,1] by descending order indicating the probabilities of exceedance (while the ascending order corresponds to the cumulative distribution functions $F_{v_{sim,dry}}$ or $F_{v_{sim,wet}}$ which indicates the probabilities of non-exceedance). To identify extreme summer NHD values the same quantile 0.8 used for SPEI is considered, and the correspondent probabilities of exceedance are estimated based on $1 - F_{v_{sim,dry}}(0.8)$ and $1 - F_{v_{sim,wet}}(0.8)$.

The associated uncertainties of the estimated probabilities of exceedance are addressed in terms of the uncertainty associated to the copula parameter and using the theoretical values inferred from the copula functions based on Equation 2 and Equation 3 where $q = 0.8$ is the threshold used in both variables. In other words, the effect of the copula parameters inaccuracy is here considered in terms of the 95% confidence level and applying the formulas of conditional

probabilities (Equation 2 and Equation 3) using the lower and upper bound levels of the estimated copula parameters.

$$\Pr(v \geq q \mid u \geq q) = \frac{1 - F_v(q) - F_u(q) + C(v = q, u = q)}{1 - F_u(q)} = \frac{1 - 2q + C(v = q, u = q)}{1 - q} \quad \text{(A.2)}$$

$$\Pr(v \geq q \mid u < q) = \frac{F_u(q) - C(v = q, u = q)}{F_u(q)} = \frac{q - C(v = q, u = q)}{q} \quad \text{(A.3)}$$

The upper tail dependence λ_U for the chosen copulas will be evaluated and is obtained as the limit of Equation 2 when q tends to 1 (Hartmann, 2004). For instance in the case of Gumbel and Joe copulas, the respective parametric estimators are given by $\lambda_U = 2 - 2^{\frac{1}{\theta}}$ (Nelsen, 2006).

A.3 Results

Six drought regions with different drought characteristics were identified applying a PCA and a cluster analysis over the IP (Figure A.2, left panel): northwestern (NW), northeastern (NE), central (C), western (W), southwestern (SW) and southeastern (SE) region. This regionalization is similarly to the ones obtained by Vicente-Serrano (2006) and Russo et al. (2015) and summarizes the IP's drought spatial heterogeneity including the 3-, 6- and 9-months SPEI timescales. In this way, the dependence between SPEI and NHD is here characterized over each drought region, in order to identify spatial patterns of drought influence on following summer (July and August) hot extremes.

Firstly, the relationship between SPEI at 3-, 6- and 9-months' time scales in May, June and July and the NHD in July and August over each region is examined in terms the Kendall's τ, and shown in Figure A.2. Note that the positive values are due to the use of mirrored SPEI values for simplicity, such that drought/wet conditions correspond to positive/negative values of SPEI. The dashed bar lines identify the regions that failed the dependence test (no rejection of null hypothesis H_0: $\tau = 0$), supporting that τ between SPEI and NHD is not significantly different from 0 at the significance level 0.1. Only the two southern regions (SW and SE) and the two

northern regions (NW and NE) show no significant dependency between NHD and SPEI in rare cases when drought is characterized in May and June. In contrast, in July all the regions show significantly dependent values between NHD and SPEI, and the C and W regions show significant dependencies in all months and timescales.

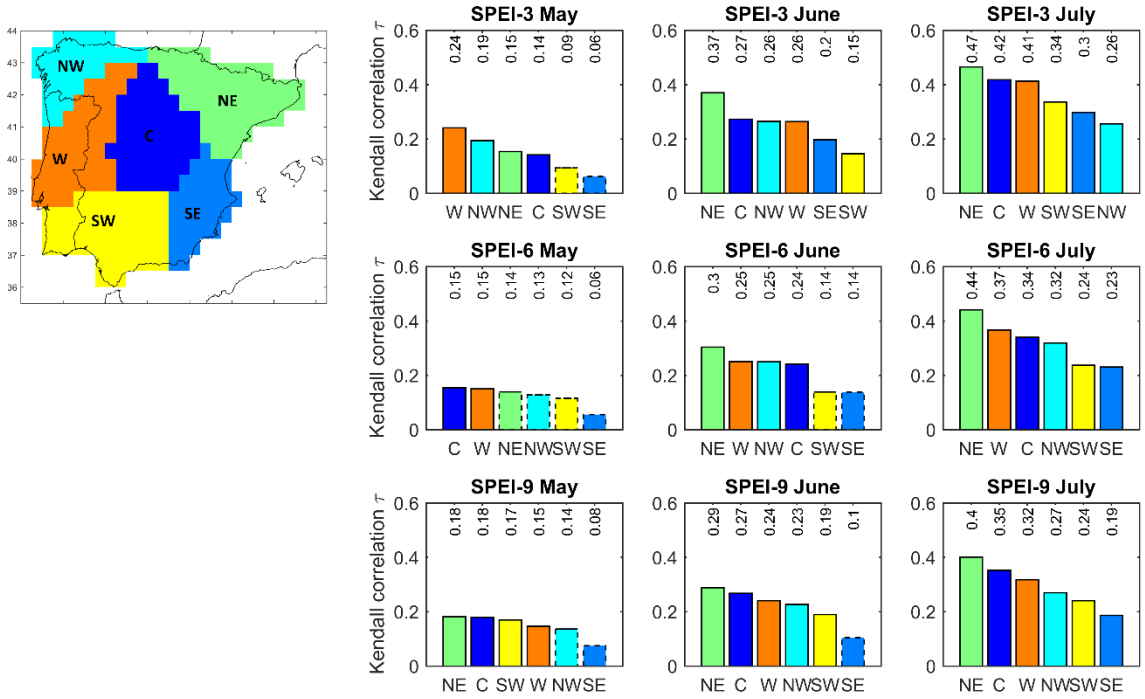


Figure A.2 - Iberian Peninsula drought regions on the left map (northwestern (NW), northeastern (NE), central (C), western (W), southwestern (SW) and southeastern (SE) region) and respective Kendall's correlation coefficient (τ) between SPEI at 3-, 6- and 9-months' time scales in May, June and July and the sum of NHD in July and August for the 1950-2014 period over each region. The bars are displayed in descending order of τ values and the dashed lines indicate the regions that failed the dependence test (p-value > 0.1) based on the Kendall's τ.

Figure A.2 illustrates the regional dependence between spring and early summer SPEI and summer NHD, as obtained by Kendall's τ. In general, dependence increases with the co-occurrence in time between SPEI and NHD in all regions, since τ increases from May to July. In contrast, the differences between different SPEI timescales are not so evident, but the shorter timescale (3-months) are generally more dependent (but not in the same regions). The NE exhibits the strongest τ in almost all months and timescales (with the exception of SPEI-3 and -6 May), in contrast to the SE region that exhibits the lowest τ in almost all months and timescales (with the exception of SPEI-3 June and July, but still low). This pattern may be related with a strong vertical gradient in the eastern part of the IP and the importance of a regional analysis.

To all possible combinations between spring SPEI and following summer NHD, six copula families (Normal (N), t, Clayton (C), Gumbel (G), Frank (F) and Joe (J)) were tested and the selected models according to the BIC values are summarized in Table A.2. In general, most of the selected models correspond to upper tail dependent copulas (Gumbel (G) and Joe (J) copulas), with the exception of the NW region which is mainly characterized by Normal copulas (Frank copulas are also selected in rare cases in NW and W regions). For verification purposes, Table A.3 summarizes the results of the parametric bootstrapping method to evaluate the goodness-of-fit of the selected models. According to Table A.3 all the p-values of the test with null hypothesis $H_0: C_\theta \in C_n$ (C_θ is the selected parametric copula and C_n the respective empirical copula) are greater than a significance level 0.1, suggesting the adequacy of the selected copulas.

Table A.2 - Copula models (Normal (N), Clayton (C), Frank (F), Gumbel (G), Joe (J)), selected based on the Bayesian Information Criteria (BIC) to characterize the joint behavior between SPEI at 3-, 6- and 9-months' time scales in May, June and July and the sum of NHD in July and August for the 1950-2014 period in the Iberian Peninsula at the regional level (NW, NE, C, W, SW and SE region).

		IB regions											
		C		**SE**		**NW**		**NE**		**SW**		**W**	
		Model	λ_U	Model	λ_U	Model	λ_U	Model	λ_U	Model	λ_U	Model	λ_U
	May	J	0.31	N	-	N	-	N	-	J	0.30	J	0.43
SPEI	June	J	0.49	J	0.42	N	-	G	0.45	J	0.32	G	0.36
3-month	July	G	0.54	J	0.55	N	-	G	0.58	N	-	N	-
	May	J	0.33	N	-	N	-	N	-	G	0.19	J	0.32
SPEI	June	J	0.47	J	0.33	N	-	N	-	J	0.33	J	0.45
6-month	July	G	0.45	J	0.42	N	-	N	-	J	0.46	G	0.46
	May	N	-	N	-	F	-	N	-	G	0.24	J	0.29
SPEI	June	J	0.48	J	0.29	N	-	N	-	G	0.26	F	-
9-month	July	N	-	J	0.36	N	-	N	-	G	0.33	F	-

Table A.3 - p-value of the copula models selected based on the Bayesian Information Criteria (BIC) to characterize the joint behavior between SPEI at 3-, 6- and 9-months' time scales in May, June and July and the sum of NHD in July and August for the 1950-2014 period in the Iberian Peninsula at the regional level (NW, NE, C, W, SW and SE region).

		IB regions					
		C	**SE**	**NW**	**NE**	**SW**	**W**
	May	0.30	0.21	0.38	0.38	0.87	0.55
SPEI	June	0.62	0.57	0.79	0.35	0.58	0.66
3-month	July	0.35	0.67	0.76	0.73	0.19	0.54
	May	0.30	0.62	0.17	0.55	0.63	0.54
SPEI	June	0.23	0.48	0.67	0.26	0.41	0.56

6-month	July	0.49	0.47	0.74	0.61	0.28	0.58
	May	0.14	0.35	0.19	0.45	0.51	0.42
SPEI	June	0.29	0.62	0.26	0.27	0.64	0.12
9-month	July	0.14	0.56	0.85	0.16	0.23	0.30

According to Table A.2, a spatial pattern of the joint behavior between spring and early summer SPEI and summer NHD is characterized based on the type of selected copulas: most of the regions are dominated by asymmetrical dependence structures with upper tail dependence, while the NW is only characterized by symmetric dependence structures (Table A.2). In terms of timescales, the longest one (9-month) is mainly characterized by Normal (N) or Frank (F) copulas ($\lambda_U = 0$), which can be explained by the information of the previous 9-months that can have less influence on the following summer hot extremes. In contrast, shorter timescales (3- and 6-months) are mainly described by Gumbel (G) or Joe (J) copulas, given that the drought conditions closest to the summer months may exert strongest influence in the following hot extremes. In fact, λ_U values are generally higher in June and July at both 3- and 6-months timescales.

Some features in Table A.2 may contrast with Figure A.2. Even though the NE exhibits the strongest τ (Figure A.2) in most of the cases, in terms of λ_U is only more accentuated considering SPEI with the shorter timescale. In fact, the relationship between summer NHD in the NE region and SPEI in June and July at 3-month timescale is described by a Gumbel (G) model (extreme value copula), but the other months and timescales of SPEI in the NE feature Normal (N) copulas. In addition, the pointed out vertical gradient in the eastern part (NE and SE) of the IP based on Kendall's τ (Figure A.2), is not so evident in terms of upper tail dependence at the 3-month timescale, which show similar λ_U values in NE and SE. Another feature in Table A.2 contrasting Figure A.2, is in the SW region, which displayed rather modest Kendall's τ values but is the region with the higher number of Gumbel (G) models describing the relationship between SPEI and NHD. Still, in comparison to the other regions, a larger λ_U in the SW is only observed in one case, with July with a SPEI-6 (Table A.2).

In order to understand if summer hot extremes are exacerbated by the previous dryness, samples of SPEI and NHD generated from the selected copulas are further used to compare the summer NHD behavior when preceded by dry conditions ($v_{sim,dry}$) or by normal/wet conditions ($v_{sim,wet}$). Figure A.3 and Figure A.4 show the conditional survival functions of NHD under dry and normal/wet conditions, respectively. The probabilities of exceedance when preceded

by dry conditions (Figure A.3) display a much larger spatial heterogeneity between regions in comparison to when preceded by normal/wet conditions (Figure A.4). Moreover, the survival curves in Figure A.3 are always more accentuated than survival curves in Figure A.4, suggesting that the probabilities of exceedance are always higher when NHD is preceded by dryness rather than by normal/wet conditions.

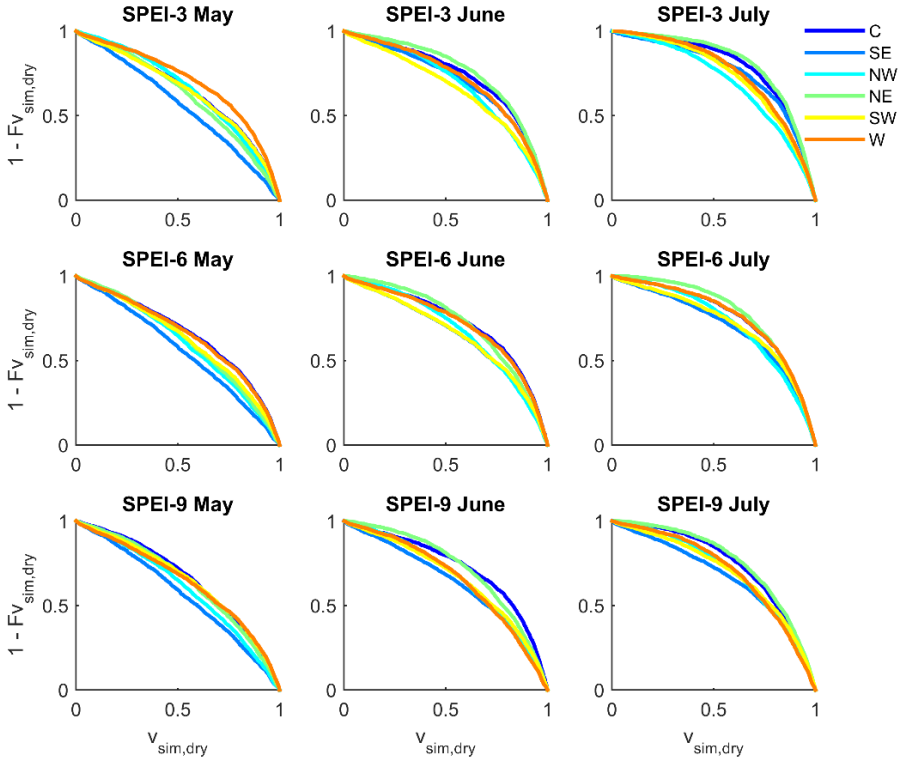


Figure A.3 - Conditional survival curves based on the samples of summer NHD generated from the selected copulas (Table A. 2) under dry conditions ($v_{sim,dry}$), indicating the exceedance probability $1 - F_{v_{sim,dry}}$ with SPEI at 3-, 6- and 9-months' time scales in May, June and July for each drought region in the Iberian Peninsula at the regional level (NW, NE, C, W, SW and SE region).

According to Figure A.3, the regions with more accentuated survival curves (higher NHD probabilities of exceedance) are the W, NE and C, depending on the month and timescale of SPEI. In the case of W region, the NHD survival curves conditioned by drought conditions in May at all timescales are mainly on the top, which suggest that summer extreme temperatures in the W are mainly conditioned by spring drought. This is in agreement with the values of λ_U in Table A.2, which point out a littoral influence of drought conditions in May as well. Similarly, in June the region with strongest λ_U is the C in all timescales, and the lead of C survival curves in June is also notable at 6- and 9-months. In contrast, the southern regions (SE and SW) and in some cases the NW region, are mainly associated to the lower survival curves, suggesting lower probabilities of exceedance (Figure A.3). Generally, the survivals curves are

more accentuated from May to July (like Figure A.2), and the differences between timescales is not so evident (Figure A.3). However, SPEI-3 in May and July may display larger differences between regions, while the curves in the other cases are closest between regions.

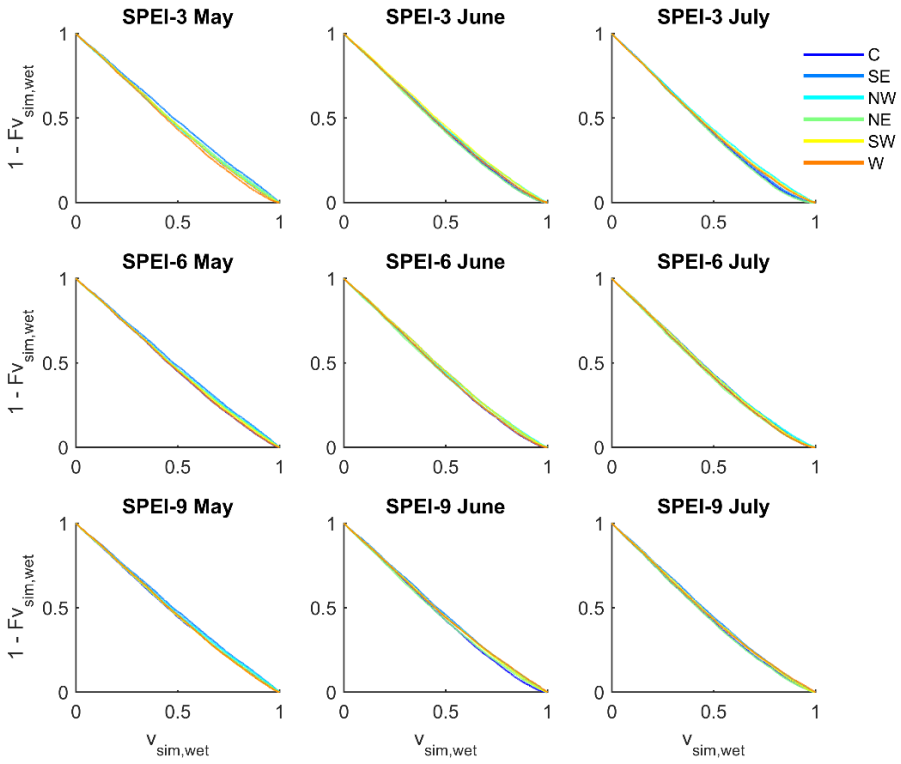


Figure A.4 - Same as Figure A.3 for wet/normal conditions.

Based on the conditional survival functions (Figure A.3 and Figure A.4), the conditional probability of NHD exceeding the quantile $q = 0.8$ (the same threshold used here to define drought conditions) over each region when preceded by dry or normal/wet conditions is obtained by $1 - F_{v_{sim,dry}}(0.8)$ and $1 - F_{v_{sim,wet}}(0.8)$ and shown in Figure A.5 and Figure A.6, respectively. Similarly to Figure A.2, the regions are displayed in decreasing order of conditional probability of exceedance in Figure A.5 and Figure A.6. As expected from the previous shown survival curves (Figure A.3 and Figure A.4), the conditional probabilities are always higher when NHD is preceded by dryness (Figure A.5) rather than by normal/wet conditions (Figure A.6). Moreover, Figure A.5 displays a much larger spatial heterogeneity in comparison to Figure A.6, showing that the response of NHD extremes to dryness varies between regions over the IP territory (in contrast, the results suggest that the response of NHD extremes to wetness is similar between regions). This result suggests that water surplus do not influence as much as water stress influences summer extreme temperatures (in agreement with

the fact that most of the regions feature models capturing the behavior in the upper extreme quantiles), and that the drought-related hot extremes varies significantly among regions in contrast to Figure A.6. In general, the probabilities of exceedance hot extremes at the $q = 0.8$ under dry conditions exceed 40% in most of the cases (except some regions in May and June 9-month), and under wet regimes the same chances are lesser than 19% in all cases.

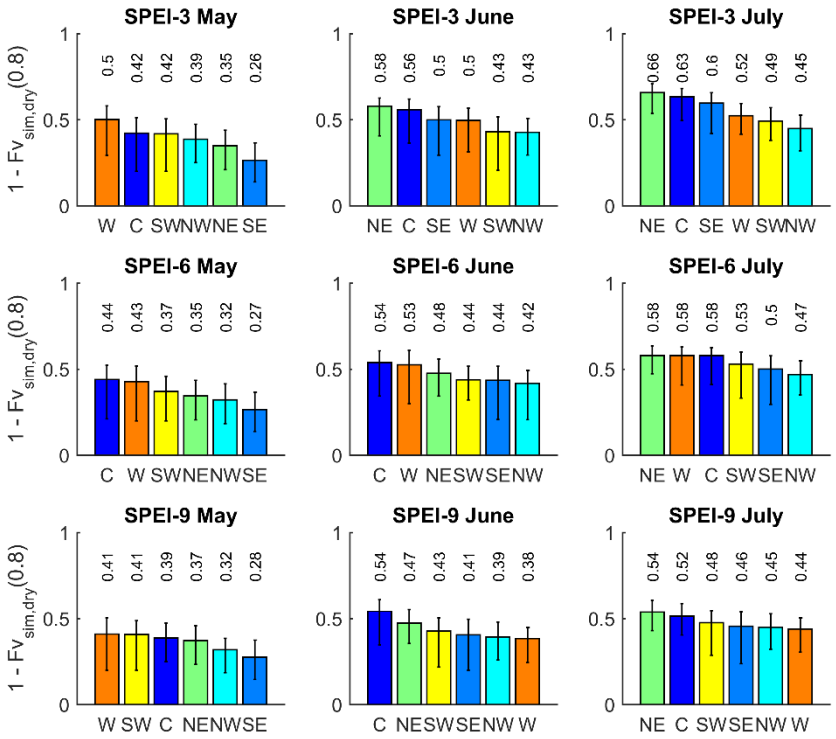


Figure A.5 - Conditional probability of summer NHD exceeding the quantile $q = 0.8$ based on the copula simulations over each drought region when preceded by dryness ($1 - F_{v_{sim,dry}}(0.8)$) with SPEI at 3-, 6- and 9-months' time scales in May, June and July for each drought region in the Iberian Peninsula at the regional level (NW, NE, C, W, SW and SE region).

According to Figure A.5, the regions with highest conditional probabilities of exceedance of hot extremes at the $q = 0.8$ when preceded by dryness are the W, NE and C regions (also in accordance with the survival functions in Figure A.3). When July is a dry month, NE is the region with highest probability of occurring summer hot extremes at all timescales, whereas when June is a dry month, is slightly more likely of occurring summer hot extremes induced by dryness at 6- and 9-months timescales in the C region (Figure A.5). If May is dry, the W region is where most of the following summer hot extremes may be induced by dryness as shown at all timescales (as previously shown in Figure A.3), despite being the second in raking in case of 6 months, but very close to the first in the ranking.

When May is a dry month, the SE region is the one with lowest probabilities of occurring hot extremes at all timescales (Figure A.5), while in June and July, NW is the region with lowest probabilities of occurring hot extremes in the following summer at the shorter timescales (3- and 6-months). At the longest timescale (9-months), the W region is the one with less likelihood of having hot extremes induced by drought conditions (in contrast with a dry May in the same region). Hence, the way how water stress affects the following summer hot extremes varies with the month, timescale of event and region (Figure A.5). The same conclusion is not so evident when the hottest months are preceded by water surplus or normal conditions (Figure A.6).

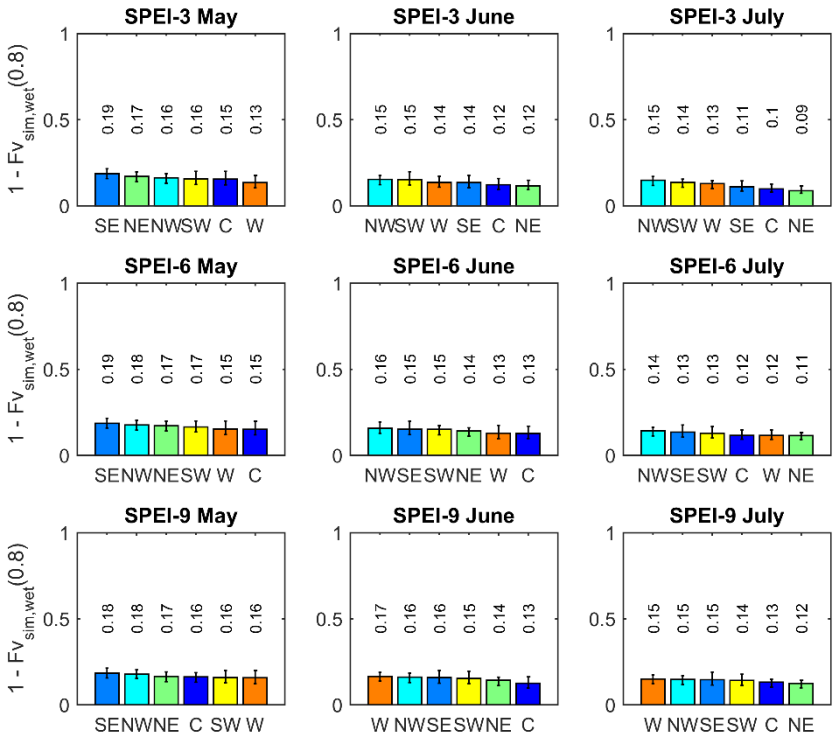


Figure A.6 - Same as Figure A.5 for wet/normal conditions.

As conditional probabilities in Figure A.6 do not vary as much between months, timescales and regions as Figure A.5, bar graphs illustrating the differences between dryness and wetness would have a similar shape to Figure A.5. As so, the transition from previous wet to dry regimes has the biggest impact in July, followed by June and May. In the same way, the regions with largest changes in conditional probabilities from wet to dry regimes are NE, C and W regions, depending on the month and timescale. In the case of the NE region in July at 3-month timescale the conditional probabilities of hot extremes may increase 57% when we transit from water surplus to deficits. In contrast, in the SE region in May at all timescales, the changes between

Figure A.5 and 7 are only around 10%, because it is the region with highest likelihood of exceeding extreme values of NHD after soil moisture surplus in this case.

A.4 Discussion

The likelihood of drought-related hot extremes in the summer were evaluated considering the NHD during July and August and the preceding 3, 6- and 9-month SPEI during May, June and July, over the IP at the regional level. The presented results intended to address the following key questions:

1) Is previous dryness associated to extreme temperatures in the hottest months of the IP at the regional level?

2) Is it possible to identify a spatial pattern of the dependence structures between spring and early summer SPEI and summer NHD?

3) What is the probability of occurrence of extreme summer hot days over each region when preceded by drought conditions?

The crucial importance of addressing these questions in the IP is related to the consequences of preceding lack of precipitation and the associated depletion of soil moisture which reduces the evaporative cooling and, thereby increases atmospheric heating and sensible heat flux amplifying the summer temperature extremes (Fischer et al., 2007; Mueller and Seneviratne, 2012). In addition, drought-related hot extremes have been closely associated to outstanding fires such as the case of the 2003 major heatwave in Europe with a considerable proportion of burnt area in the IP (Trigo et al., 2006), followed by the severe drought of 2004/2005 also with major extent of burnt area in the IP (Gouveia et al., 2012). As also shown by Gouveia et al. (2016), the Greece extreme fire season in 2007 was particularly devastating due to the combined effect of droughts and heatwaves. In this way, this work intends to adopt a compound event perspective by acknowledging the co-occurrence or sequential occurrence of dry and hot extremes, due to the importance of the feedback mechanisms involved in the soil moisture–temperature interactions and the subsequent extreme consequences (Seneviratne et al., 2010; Miralles et al., 2019).

The IP's drought spatial heterogeneity was a main factor when characterizing the water stress influence on following summer extreme temperatures (Figure A.2). Figures A.2-A.6 and Table

A.2 highlight how the dependence between SPEI and NHD varies among regions (even in the case of the upper tail dependence), particularly when we focus only on the dry conditions (Figure A.3 and Figure A5). In contrast, after normal/wet conditions the behavior between SPEI and NHD is similar between regions (Figure A.4 and Figure A.6) and the chances of hot prone July and August are less probable when compared to Figure A.3 and Figure A.5.

In addition, although most of the relationships measured based on τ, λ_U and conditional probabilities describe a general increase from May towards July, the regional increasing order is not the same among different measures. In other words, the regions with strongest τ (Figure A.2), not always correspond to the regions with strongest λ_U (Table A.2) or conditional probabilities of exceedance (Figure A.3 and Figure A.5). In the same way, regions exhibiting lower τ between SPEI and NHD like the SW region, show $\lambda_U \neq 0$, suggesting how variables with no apparent noteworthy correlation may show tail dependence in extreme values. Moreover, while NE and SE regions have similar λ_U values, the τ values pointed out vertical gradient in the relationship between NHD and SPEI over the eastern part of the IP. Note that the NE region in July with SPEI-3 displays the highest τ (Figure A.2), λ_U (Table A.2) and conditional probability of exceedance under dry conditions (Figure A.3 and Figure A.5), but not the highest conditional probability of exceedance preceded by wetness (Figure A.6), suggesting that is the upper extreme dependence that controls this relationship the most. On the other hand, although the W region in May SPEI-3 shows a modest $\tau = 0$, it's the region with highest λ_U and probability of exceedance at May SPEI-9, being the region with highest predictive power of summer hot extremes. Followed by NE, the C region is also one of the most summer hot prone areas (Table A.2, Figure A.3 and Figure A.5). In contrast, the SE region exhibited the lowest τ in almost all months and timescales, and it's the region with less difference in the conditional probabilities when we change from a wet to a dry regime (Figure A.5 and Figure A.6). This finding suggest that other trigger(s) rather than water stress is(are) inducing hot extremes in the SE region of the IP.

According to the results, the influence of July dry conditions on summer hot extremes is more evident in NE regions, while May and June drought may influence more the number of warmer days in W and C regions, respectively. Moreover, the strength of this relation progresses from spring to summer and, subsequently, from W to C and NE (varying with timescale). Although the use of different methodologies makes difficult a direct comparison between studies, this conclusion agrees with previous works that found relationships between very dry conditions in

spring and the evolution of warm days in W and C regions (Fernández-Montes et al., 2013). Additionally, the dominance of extreme temperatures in the NE part of the IP is also acknowledged in several studies (Díaz et al., 2006; Rodríguez-Puebla et al., 2010; El Kenawy et al., 2012; Fernández-Montes et al., 2013), including during the 2003 heatwave (Díaz et al., 2006). The NE of the IP includes the Ebro valley, which is a semi-arid region with increasing drought severity (Vicente-Serrano and Cuadrat-Prats, 2007; Vicente-Serrano et al., 2014; Ojeda et al., 2016) and is experiencing an overall warming (El Kenawy et al., 2012). In contrast, in wet zones such as the NW where drought is not a major limiting factor (Ojeda et al., 2016) we do not identify a key influence on the occurrence of hot summers.

Other relevant mechanisms before and during heatwaves, such as the associated atmospheric circulation, was beyond the present study's scope, and it has been assessed by other authors which discuss the role played blocking patterns and ridge of high pressure (Tomczyk et al., 2017; Sousa et al., 2018). More recently, Sousa et al. (2019) have discussed the influence of Saharan air intrusions as a relevant mechanism for the occurrence of hot extremes in the IP. Other studies focus on the synoptic diagnosis of both heatwaves and droughts (Fink et al., 2006).

Overall, the results suggest that when SPEI on previous months indicates dryness, rather than normal/wet conditions, the probabilities of occurring hot extremes during the following hottest months rises substantially. This result is in accordance with previous works, which identified in the IP significantly correlations between NHD and drought indicators, i.e., high values of NHD follows SPI/SPEI values associated to drought (Mueller and Seneviratne, 2012; Russo et al., 2019). Although these works pinpoint that there is a strong association and a certain predictability between preceding drought conditions and summer extreme temperatures, they lack a joint probability approach to address the risk of surpassing certain threshold conditions. The present approach has gone one step further by adopting a copula-based methodology which allowed for the estimation of conditional probabilities for specific thresholds of SPEI and NHD.

In this work we have modeled the dependence structures between a drought indicator and following extreme temperatures by fitting Archimedean and Elliptical copula functions to estimate their joint probability. While previous similar works have analyzed the dependence between dry and hot conditions considering only normal copulas (Hao et al., 2017, 2018b) or correlation analysis (Russo et al., 2019), we found that this combined phenomenon is very well

identified for the most of the IP's regions by asymmetrical copulas with upper tail dependence (except in NW regions as shown in Table A.2).

One of the most important outcomes of this study is the proven ability to estimate the conditional probability of drought-related hot extremes in the summer of the IP for different regions. Similarly, the study by Mueller and Seneviratne (2012) have determined the occurrence probability for above-average NHD for the globe, and found that after wet conditions is very unlikely to exceed above-average NHD in most of the areas of the world, including the IP. In this study, we have considered more extreme thresholds of NHD and estimated the conditional probability of summer NHD exceeding the quantile $q = 0.8$ given antecedent drought conditions. Moreover, the conditional probabilities of exceedance determined in the present study were estimated based on copula-simulations of the involved SPEI and NHD, thus allowing to estimate drought-related hot extremes likelihood with data preserving the dry-hot dependence structures based on their joint probability.

Regarding the European 2003 and 2010 extreme heatwaves concurrent with serious drought conditions, the study by Barriopedro et al. (2011) have assessed their return periods based on regional multi-model experiments and have found 2003-analogs before 2050 in more than half of the models, while 2010-analogs only feature after 2050. This means that the likelihood of 2010-like events are low until the second half of the 21st century, but 2003-like events could be much likely on the next 30 years. In the present study, we haven't focused on the conditional probabilities of exceedance NHD values of the 2003 and 2010 events as they are such extreme events that they exceed the 0.8 quantile. Moreover, other recent summers, such as 2015 and 2018, did not feature European mega heatwaves, but did feature regional extreme events (Sousa et al., 2019), and for these reason we have prioritized to perform the present study at the regional scale of the IP. As a matter of fact, the heterogeneity found in the drought spatial patterns (Figure A.2) and subsequent relationships with following summer hot extremes (Figure A.3 and Figure A.5) highlights the regional character of hot and dry relationships.

This study provides estimates of the probability of drought-related hot summers over the IP at the regional level, which could be of high importance for responsible authorities to be able to reduce the cost of the associated impacts. The proven predictive capacity of drought-related hot summers may provide some guidance and help to contribute to design prevention measures to

mitigate the impacts magnified by the interactions between the different hazards (human health impacts, fire risk, air pollution, crop-loss, etc.).

A.5 Conclusions

The present study aimed to investigate the conditional probability of compound dry and hot extremes in the IP at the regional level, particularly summer hotness induced by previous dryness. The main findings of this study are summarized below:

- The dependence structure between SPEI and NHD is very well identified for the most of the IP's regions by asymmetrical copulas with upper tail dependence (except in NW regions), suggesting that compound hot and dry extremes are strongly associated.
- The transition from previous wet to dry regimes increases substantially the probability of exceeding NHD extreme values suggesting how summer hot extremes may be induced by previous soil moisture deficits.
- The relationships between summer NHD and previous SPEI increase from May to July, suggesting that the predictability power of a hot prone July and August increases as summer arrives.
- IP's drought spatial heterogeneity was found to be a main factor when characterizing the water stress influence on following summer extreme temperatures.
- In general, NE, W and C regions were found to be the regions with highest conditional probabilities of exceedance of hot extremes when preceded by dryness, in contrast to SW, NW and SE regions (depending on the SPEI month and timescale).

Future assessments of the amplified impacts of this kind of compound events in comparison with the impacts of individual drought or hotness is one of the targets for further work. Namely the impacts of compound drought and heat on the agricultural crops and how to prevent associated losses (Feng et al., 2019). Moreover, the combination of deficit precipitation and hot conditions may also exacerbate the fire risk and air pollution and understand the changes in these likelihoods are other promising lines for future research in the present climate chance context. Nonetheless, the development of such studies requires higher dimensional models with more complex structures and other methods rather than bivariate copulas.

Supplementary Material

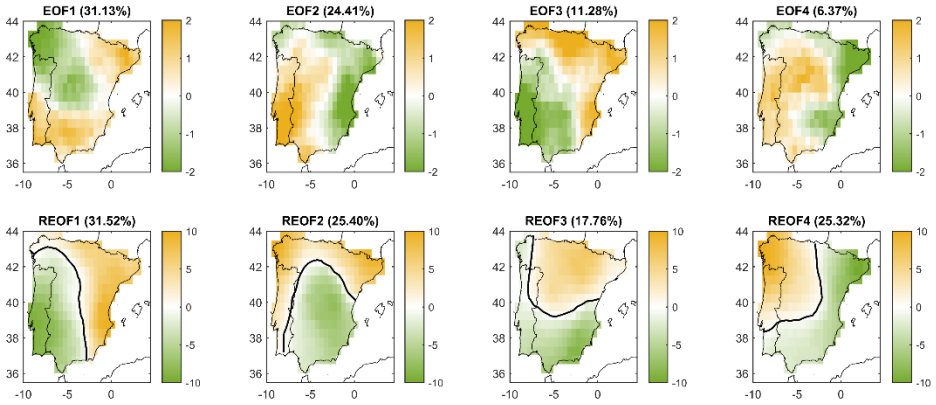


Figure SA.1 - Top: Four main Principal Component Analysis (PCA) modes considering the SPEI during May and June at the 3-month timescale. Bottom: Varimax rotated PCA modes, from which the main 3 were retained.

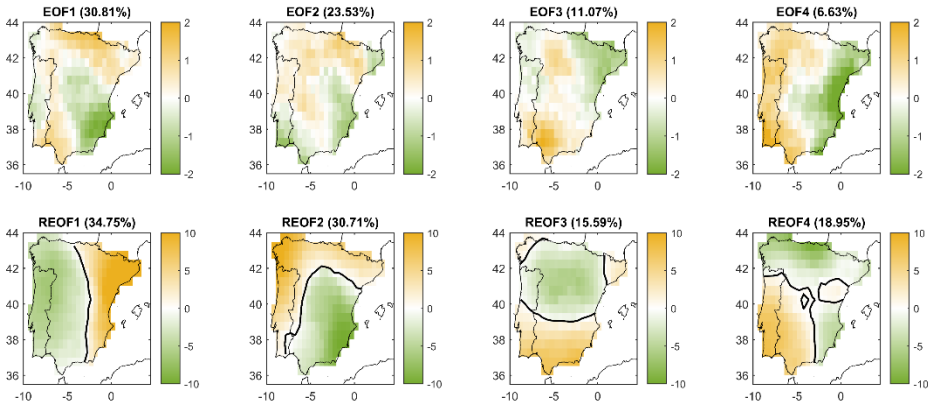


Figure SA.2 - Same as Fig. S1 for the 6-month timescale.

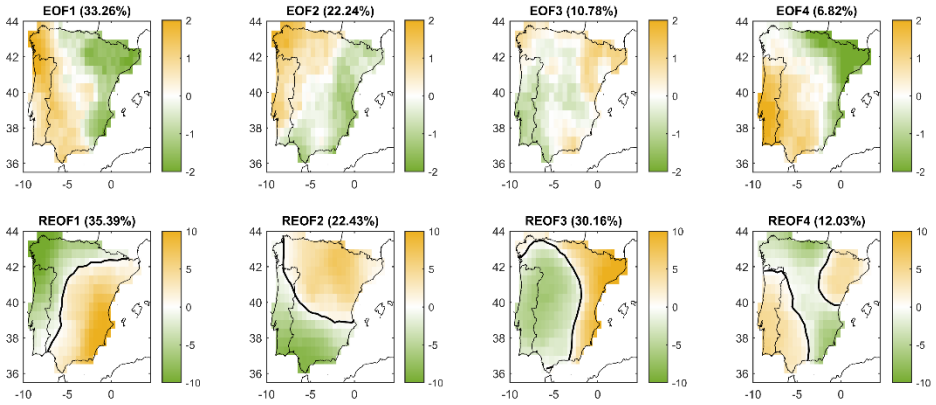


Figure SA.3 - Same as Fig. S1 for the 9-month timescale.

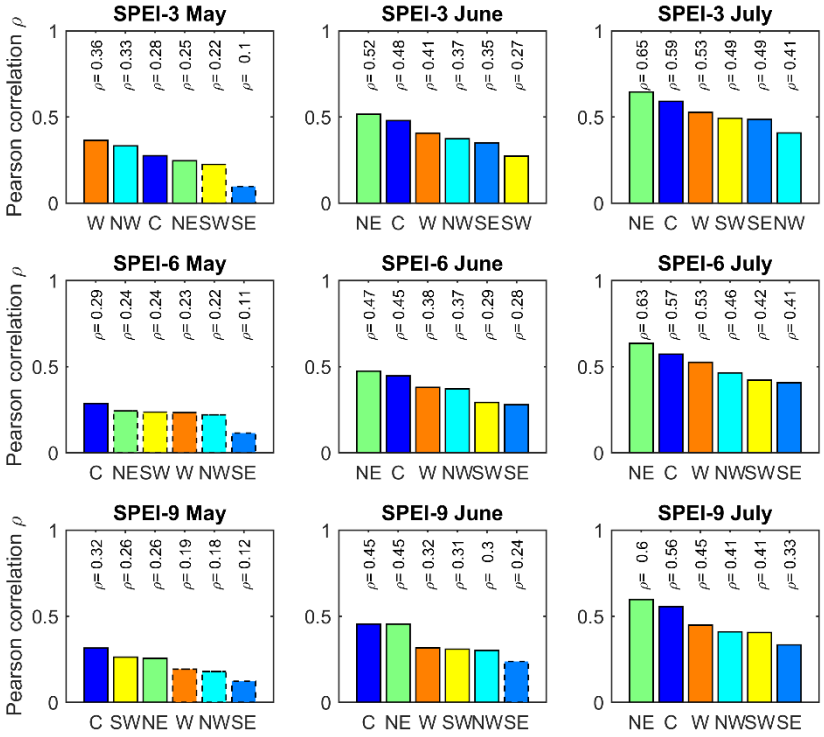


Figure SA.4 - Pearson correlation coefficient (ρ) between SPEI at 3-, 6- and 9-months' time scales in May, June and July and the sum of NHD in July and August for the 1950-2014 period over each drought region. The bars are displayed in descending order of ρ values and the dashed lines indicate the regions that p-value > 0.1.

Table SA.1 - Copula models selected based on the Akaike Information Criteria (AIC) to characterize the joint behavior between SPEI at 3-, 6- and 9-months' time scales in May, June and July and the sum of NHD in July and August for the 1950-2014 period in the Iberian Peninsula at the regional level (northwestern (NW), northeastern (NE), central (C), western (W), southwestern (SW) and southeastern (SE) region).

		IB regions					
		C	**SE**	**NW**	**NE**	**SW**	**W**
SPEI 3-month	May	Joe	Gaussian	Gaussian	Gaussian	Joe	Joe
	June	Joe	Joe	Gaussian	t	Joe	Gumbel
	July	t	Joe	Gaussian	Gumbel	Gaussian	Gaussian
SPEI 6-month	May	Joe	Gaussian	Gaussian	Gaussian	Gumbel	Joe
	June	Joe	Joe	Gaussian	Gaussian	Joe	Joe
	July	Gumbel	Joe	Gaussian	Gaussian	Joe	Gumbel
SPEI 9-month	May	Gaussian	Gaussian	Frank	Gaussian	Gumbel	Joe
	June	Joe	Joe	Gaussian	Gaussian	Gumbel	Frank
	July	Gaussian	Joe	Gaussian	Gaussian	Gumbel	Frank

References

Afshar, M. H., Sorman, A. U. and Yilmaz, M. T.: Conditional copula-based spatial-temporal drought characteristics analysis-A case study over Turkey, Water (Switzerland), 8(10), 1–16, doi:10.3390/w8100426, 2016.

Agnew, C. T.: Using the SPI to Identify Drought, Drought Netw. News, 12(2000), 6–12, 2000.

Aitsi-Selmi, A., Egawa, S., Sasaki, H., Wannous, C. and Murray, V.: The Sendai Framework for Disaster Risk Reduction: Renewing the Global Commitment to People's Resilience, Health, and Well-being, Int. J. Disaster Risk Sci., 6(2), 164–176, doi:10.1007/s13753-015-0050-9, 2015.

Alonso, C., Gouveia, C. M., Russo, A. and Páscoa, P.: Crops' exposure, sensitivity and adaptive capacity to drought occurrence, Nat. Hazards Earth Syst. Sci., 19(12), 2727–2743, doi:10.5194/nhess-19-2727-2019, 2019.

Anderson, W. B., Seager, R., Baethgen, W., Cane, M. and You, L.: Synchronous crop failures and climate-forced production variability, Sci. Adv., 5(7), 1–10, doi:10.1126/sciadv.aaw1976, 2019.

Andrade, C. and Belo-Pereira, M.: Assessment of droughts in the Iberian Peninsula using the WASP-Index, Atmos. Sci. Lett., 16(3), 208–218, doi:10.1002/asl2.542, 2015.

Atzberger, C., Formaggio, A. R., Shimabukuro, Y. E., Udelhoven, T., Mattiuzzi, M., Sanchez, G. A., Arai, E., Formaggio, A. R., Shimabukuro, Y. E. and Udelhoven, T.: Obtaining crop-specific time profiles of NDVI : the use of unmixing approaches for serving the continuity between SPOT-VGT and PROBA-V time series, Int. J. Remote Sens., 35(7), 2615–2638, doi:10.1080/01431161.2014.883106, 2014.

Austin, R. B., Cantero-Martínez, C., Arrúe, J. L., Playán, E. and Cano-Marcellán, P.: Yield-rainfall relationships in cereal cropping systems in the Ebro river valley of Spain, Eur. J. Agron., 8(3–4), 239–248, doi:10.1016/S1161-0301(97)00063-4, 1998.

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. and García-Herrera, R.: The hot summer of 2010: Redrawing the temperature record map of Europe, Science (80-.)., 332(6026), 220–224, doi:10.1126/science.1201224, 2011.

Bastos, A., Gouveia, C. M., Trigo, R. M. and Running, S. W.: Analysing the spatio-temporal impacts of the 2003 and 2010 extreme heatwaves on plant productivity in Europe, Biogeosciences, 11(13), 3421–3435, doi:10.5194/bg-11-3421-2014, 2014.

Beguería, S., Vicente-Serrano, S. M., Reig, F. and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, Int. J. Climatol., 34(10), 3001–3023, doi:10.1002/joc.3887, 2014.

Belo-Pereira, M., Dutra, E. and Viterbo, P.: Evaluation of global precipitation data sets over the Iberian Peninsula, J. Geophys. Res. Atmos., 116(20), 1–16, doi:10.1029/2010JD015481, 2011.

Ben-Ari, T., Adrian, J., Klein, T., Calanca, P., Van der Velde, M. and Makowski, D.: Identifying indicators for extreme wheat and maize yield losses, Agric. For. Meteorol., 220, 130–140, doi:10.1016/j.agrformet.2016.01.009, 2016.

Ben-Ari, T., Boé, J., Ciais, P., Lecerf, R., Van Der Velde, M. and Makowski, D.: Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France, Nat. Commun., 9(1), doi:10.1038/s41467-018-04087-x, 2018.

Bento, V. A., Gouveia, C. M., DaCamara, C. C. and Trigo, I. F.: A climatological assessment of drought impact on vegetation health index, Agric. For. Meteorol., 259(February), 286–295, doi:10.1016/j.agrformet.2018.05.014, 2018.

Bento, V. A., Gouveia, C. M., Dacamara, C. C., Libonati, R. and Trigo, I. F.: The roles of NDVI and Land Surface Temperature when using the Vegetation Health Index over dry regions, Glob. Planet. Change, 190(July 2019), 103198, doi:10.1016/j.gloplacha.2020.103198, 2020.

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M. and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in Ravenna (Italy), Hydrol. Earth Syst. Sci., 21(6), doi:10.5194/hess-21-2701-2017, 2017.

Bezak, N. and Brilly, M.: Applicability of copula functions in analysis of extreme hydrological, in Proceedings of the Mediterranean Meeting on ″Monitoring, modelling and early warning of extreme events triggered by heavy rainfalls″. PON 01_01503 - MED-FRIEND project University of Calabria, Cosenza (Italy), June 26th-28th, 2014., 2014.

Blauhut, V., Stahl, K., Stagge, J. H., Tallaksen, L. M., Stefano, L. De and Vogt, J.: Estimating

drought risk across Europe from reported drought impacts, drought indices, and vulnerability factors, Hydrol. Earth Syst. Sci., 20(7), 2779–2800, doi:10.5194/hess-20-2779-2016, 2016.

Bokusheva, R.: Improving the Effectiveness of Weather-based Insurance: An Application of Copula Approach, MPRA Pap., (62339), doi:https://mpra.ub.uni-muenchen.de/62339/, 2014.

Bokusheva, R. and Breustedt, G.: The effectiveness of weather-based index insurance and area-yield crop insurance: How reliable are ex post predictions for yield risk reduction?, Q. J. Int. Agric., 51(2), 135–156, 2012.

Bokusheva, R., Kogan, F., Vitkovskaya, I., Conradt, S. and Batyrbayeva, M.: Satellite-based vegetation health indices as a criteria for insuring against drought-related yield losses, Agric. For. Meteorol., 220, 200–206, doi:10.1016/j.agrformet.2015.12.066, 2016.

Buras, A., Rammig, A. and S. Zang, C.: Quantifying impacts of the 2018 drought on European ecosystems in comparison to 2003, Biogeosciences, 17(6), 1655–1672, doi:10.5194/bg-17-1655-2020, 2020.

Capa-Morocho, M., Ines, A. V. M., Baethgen, W. E., Rodríguez-Fonseca, B., Han, E. and Ruiz-Ramos, M.: Crop yield outlooks in the Iberian Peninsula: Connecting seasonal climate forecasts with crop simulation models, Agric. Syst., 149, 75–87, doi:10.1016/j.agsy.2016.08.008, 2016a.

Capa-Morocho, M., Rodríguez-Fonseca, B. and Ruiz-Ramos, M.: Sea surface temperature impacts on winter cropping systems in the Iberian Peninsula, Agric. For. Meteorol., 226–227, 213–228, doi:10.1016/j.agrformet.2016.06.007, 2016b.

Carvalho, P., Azam-Ali, S. and Foulkes, M. J.: Quantifying relationships between rooting traits and water uptake under drought in Mediterranean barley and durum wheat, J. Integr. Plant Biol., 56(5), 455–469, doi:10.1111/jipb.12109, 2014.

Chen, L., Singh, V. P., Chen, L., Singh, V. P., Asce, F., Guo, S., Mishra, A. K. and Guo, J.: Drought Analysis using copulas, J. Hydrol. Eng., 18(August), 797–808, doi:10.1061/(ASCE)HE.1943-5584.0000697., 2013.

CLC2012: Copernicus Programme Land Monitoring Service, [online] Available from: https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012, 2012.

COPA-COGECA: Assessment of the impact of the heat wave and drought of the summer 2003 on agriculture and forestry, Comm. Agric. Organ. Eur. Union Gen. Comm. Agric. Coop. Eur.

Union, Brussels, 15, 2003.

Corder, G. W. and Foreman, D. I.: Nonparametric Statistics: An Introduction, in Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach, pp. 1–11, John Wiley & Sons, Inc., 2011.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M. and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, J. Geophys. Res. Atmos., 123(17), 9391–9409, doi:10.1029/2017JD028200, 2018.

Cortesi, N., Gonzalez-Hidalgo, J. C., Trigo, R. M. and Ramos, A. M.: Weather types and spatial variability of precipitation in the Iberian Peninsula, Int. J. Climatol., 34(8), 2661–2677, doi:10.1002/joc.3866, 2014.

Cossani, C. M., Slafer, G. A. and Savin, R.: Yield and biomass in wheat and barley under a range of conditions in a Mediterranean site, F. Crop. Res., 112(2–3), 205–213, doi:10.1016/j.fcr.2009.03.003, 2009.

Dalezios, N. R., Blanta, A., Spyropoulos, N. V and Tarquis, a M.: Risk identification of agricultural drought for sustainable Agroecosystems, Nat. hazards earth Syst. Sci., 14(9), 2435–2448, doi:10.5194/nhess-14-2435-2014, 2014.

Dan, Z., Guoli, W. and Huicheng, Z.: Assessment on agricultural drought risk based on variable fuzzy sets model, Chinese Geogr. Sci., 21(2), 167–175, doi:10.1007/s11769-011-0456-2, 2011.

Delcour, I., Spanoghe, P. and Uyttendaele, M.: Literature review: Impact of climate change on pesticide use, Food Res. Int., 68(October), 7–15, doi:10.1016/j.foodres.2014.09.030, 2015.

Díaz, J., García-Herrera, R., Trigo, R. M., Linares, C., Valente, M. A., De Miguel, J. M. and Hernández, E.: The impact of the summer 2003 heat wave in Iberia: How should we measure it?, Int. J. Biometeorol., 50(3), 159–166, doi:10.1007/s00484-005-0005-8, 2006.

Durante, F. and Sempi, C.: Principles of Copula Theory, Taylor & F., 2015.

Estes, L. D., Bradley, B. A., Beukes, H., Hole, D. G., Lau, M., Oppenheimer, M. G., Schulze, R., Tadross, M. A. and Turner, W. R.: Comparing mechanistic and empirical model projections of crop suitability and productivity: Implications for ecological forecasting, Glob. Ecol. Biogeogr., 22(8), 1007–1018, doi:10.1111/geb.12034, 2013.

Fahr, A.: Nonparametric Analysis, Int. Encycl. Commun. Res. Methods, 1–6, doi:10.1002/9781118901731.iecrm0168, 2017.

Falco, S. Di, Adinolfi, F., Bozzola, M. and Capitanio, F.: Crop Insurance as a Strategy for Adapting to Climate Change, J. Agric. Econ., 65(2), 485–504, doi:10.1111/1477-9552.12053, 2014.

FAO: The state of the world's land and water resources for food and agriculture, FAO (Food and Agriculture Organization of the United Nations)., 2011.

FAO: Statistical Yearbook 2014: Europe and Central Asia food and agriculture, FAO (Food and Agriculture Organization of the United Nations). [online] Available from: http://www.bfi.org.uk/statisticalyearbook2014/, 2014.

FAO: FAO Statistical Pocketbook, FAO (Food and Agriculture Organization of the United Nations)., 2015.

Feng, S., Hao, Z., Zhang, X. and Hao, F.: Probabilistic evaluation of the impact of compound dry-hot events on global maize yields, Sci. Total Environ., 689, 1228–1234, doi:10.1016/j.scitotenv.2019.06.373, 2019.

Fernández-Montes, S., Rodrigo, F. S., Seubert, S. and Sousa, P. M.: Spring and summer extreme temperatures in Iberia during last century in relation to circulation types, Atmos. Res., 127, 154–177, doi:10.1016/j.atmosres.2012.07.013, 2013.

Ferrise, R., Moriondo, M. and Bindi, M.: Probabilistic assessments of climate change impacts on durum wheat in the Mediterranean region, Nat. Hazards Earth Syst. Sci., 11(5), 1293–1302, doi:10.5194/nhess-11-1293-2011, 2011.

Fink, A. H., Brücher, T., Krüger, A., Leckebusch, G. C., Pinto, J. G. and Ulbrich, U.: The 2003 European summer heatwaves and drought – synoptic diagnosis and impacts, Weather, 209–216, doi:10.1256/wea.73.04, 2006.

Fischer, E. M., Seneviratne, S. I., Lüthi, D. and Schär, C.: Contribution of land-atmosphere coupling to recent European summer heat waves, Geophys. Res. Lett., 34(6), 1–6, doi:10.1029/2006GL029068, 2007.

Ganguli, P. and Reddy, M. J.: Risk Assessment of Droughts in Gujarat Using Bivariate Copulas, Water Resour. Manag., 26(11), 3301–3327, doi:10.1007/s11269-012-0073-6, 2012.

Ganguli, P. and Reddy, M. J.: Probabilistic assessment of flood risks using trivariate copulas, Theor. Appl. Climatol., 111(1–2), 341–360, doi:10.1007/s00704-012-0664-4, 2013.

Garcia-Herrera, R., Díaz, J., Trigo, R. M., Luterbacher, J. and Fischer, E. M.: A review of the european summer heat wave of 2003, Crit. Rev. Environ. Sci. Technol., 40(4), 267–306, doi:10.1080/10643380802238137, 2010.

García-Herrera, R., Hernández, E., Barriopedro, D., Paredes, D., Trigo, R. M., Trigo, I. F. and Mendes, M. A.: The Outstanding 2004/05 Drought in the Iberian Peninsula: Associated Atmospheric Circulation, J. Hydrometeorol., 8(3), 483–498, doi:10.1175/JHM578.1, 2007.

García del Moral, L. F., Rharrabti, Y., Villegas, D. and Royo, C.: Evaluation of Grain Yield and Its Components in Durum Wheat under Mediterranean Conditions: An Ontogenic Approach, Agron. J., 95, 266–274, 2003.

Gaupp, F., Hall, J., Hochrainer-stigler, S. and Dadson, S.: Changing risks of simultaneous global breadbasket failure, Nat. Clim. Chang., doi:10.1038/s41558-019-0600-z, 2019.

Gbegbelegbe, S., Chung, U., Shiferaw, B., Msangi, S. and Tesfaye, K.: Quantifying the impact of weather extremes on global food security: A spatial bio-economic approach, Weather Clim. Extrem., 4, 96–108, doi:10.1016/j.wace.2014.05.005, 2014.

Genest, C. and Remillard, B.: Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models, Ann. l'institut Henri Poincare Probab. Stat., 44(6), 1096–1127, doi:10.1214/07-AIHP148, 2008.

Genest, C., Ghoudi, K. and Rivest, L.-P.: A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions, Biometrika, 82(3), 543–552, 1995.

Gil, M., Garrido, A. and Gómez-Ramos, A.: Economic analysis of drought risk: An application for irrigated agriculture in Spain, Agric. Water Manag., 98(5), 823–833, doi:10.1016/j.agwat.2010.12.008, 2011.

Giménez, L., Petillo, M. G., Paredes, P. and Pereira, L. S.: Predicting Maize Transpiration , Water Use and Productivity for Developing Improved Supplemental Irrigation Schedules in Western Uruguay to Cope with Climate Variability, Water, 8(309), 1–22, doi:10.3390/w8070309, 2016.

Giorgi, F. and Lionello, P.: Climate change projections for the Mediterranean region, Glob. Planet. Change, 63(2), 90–104, 2008.

Gouveia, C. and Trigo, R. M.: Influence of climate variability on European agriculture-analysis of winter wheat production, geoENV VI–Geostatistics Environ. Appl. Springer Netherlands, 27, 335–345, doi:10.3354/cr027135, 2008.

Gouveia, C., Trigo, R. M., Dacamara, C. C., Libonati, R. and Pereira, J. M.: The North Atlantic Oscillation and European vegetation dynamics, Int. J. Climatol., 28(14), 1835–1847, doi:10.1002/joc.1682, 2008.

Gouveia, C., Trigo, R. M. and DaCamara, C. C.: Drought and vegetation stress monitoring in Portugal using satellite data, Nat. Hazards Earth Syst. Sci., 9(1), 185–195, 2009.

Gouveia, C., Liberato, M. L. R., DaCamara, C. C., Trigo, R. M. and A.Ramos, M.: Modelling past and future wine production in the Portuguese Douro Valley, Clim. Res., 48(2–3), 349–362, doi:10.3354/cr01006, 2011.

Gouveia, C. M., Bastos, A., Trigo, R. M. and Dacamara, C. C.: Drought impacts on vegetation in the pre- and post-fire events over Iberian Peninsula, Nat. Hazards Earth Syst. Sci., 12(10), 3123–3137, doi:10.5194/nhess-12-3123-2012, 2012.

Gouveia, C. M., Páscoa, P., Russo, A. and Trigo, R. M.: Land degradation trend assessment over Iberia during 1982-2012, Cuad. Investig. Geográfica, 42(1), 89, doi:10.18172/cig.2945, 2016a.

Gouveia, C. M., Bistinas, I., Liberato, M. L. R., Bastos, A., Koutsias, N. and Trigo, R.: The outstanding synergy between drought, heatwaves and fuel on the 2007 Southern Greece exceptional fire season, Agric. For. Meteorol., 218–219, doi:10.1016/j.agrformet.2015.11.023, 2016b.

Gouveia, C. M., Trigo, R. M., Beguería, S. and Vicente-serrano, S. M.: Drought impacts on vegetation activity in the Mediterranean region : An assessment using remote sensing data and multi-scale drought indicators, Glob. Planet. Change, 151, 15–27, doi:10.1016/j.gloplacha.2016.06.011, 2017a.

Gouveia, C. M., Trigo, R. M., Beguería, S. and Vicente-Serrano, S. M.: Drought impacts on vegetation activity in the Mediterranean region: An assessment using remote sensing data and

multi-scale drought indicators, Glob. Planet. Change, 151, doi:10.1016/j.gloplacha.2016.06.011, 2017b.

Grasso, M. and Feola, G.: Mediterranean agriculture under climate change: Adaptive capacity, adaptation, and ethics, Reg. Environ. Chang., 12(3), 607–618, doi:10.1007/s10113-011-0274-1, 2012.

Guttman, N. B.: Comparing the Palmer Drought Index and the Standardized Precipitation Index, J. Am. Water Resour. Assoc., 34(1), 113–121, doi:10.1111/j.1752-1688.1998.tb05964.x, 1998.

Hannachi, A., Jolliffe, I. T. and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, Int. J. Climatol., 27, 1119–1152, doi:10.1002/joc.1499 Review, 2007.

Hao, Z., Hao, F., Singh, V. P. and Ouyang, W.: Quantitative risk assessment of the effects of drought on extreme temperature in eastern China, J. Geophys. Res. Atmos., 122(17), 9050–9059, doi:10.1002/2017JD027030, 2017.

Hao, Z., Hao, F., Singh, V. P., Xia, Y., Shi, C. and Zhang, X.: A multivariate approach for statistical assessments of compound extremes, J. Hydrol., 565, 87–94, doi:10.1016/j.jhydrol.2018.08.025, 2018a.

Hao, Z., Hao, F., Singh, V. P. and Zhang, X.: Changes in the severity of compound drought and hot extremes over global land areas, Environ. Res. Lett., 13(12), doi:10.1088/1748-9326/aaee96, 2018b.

Hao, Z., Singh, V. P. and Hao, F.: Compound extremes in hydroclimatology: A review, Water (Switzerland), 10(6), 16–21, doi:10.3390/w10060718, 2018c.

Harris, I., Jones, P. D., Osborn, T. J. and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, Int. J. Climatol., 34(3), 623–642, doi:10.1002/joc.3711, 2014.

Harris, I., Osborn, T. J., Jones, P. and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Sci. Data, 7(1), 1–18, doi:10.1038/s41597-020-0453-3, 2020.

Hartmann, S. . S. and C. . G. . de V.: Asset Market Linkages in Crisis, Rev. Econ. Stat., 86(1),

313–326 [online] Available from: https://www.jstor.org/stable/3211675, 2004.

Hauser, M., Orth, R. and Seneviratne, S. I.: Role of soil moisture vs . recent climate change for heat waves in western Russia, Geophys. Res. Lett., 43(October), 2819–2826, doi:10.1002/2016GL068036.Received, 2015.

Haylock, M. R., Hofstra, N., Tank, A. M. G. K., Klok, E. J., Jones, P. D. and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, , 113, doi:10.1029/2008JD10201, 2008.

Hernandez-Barrera, S., Rodriguez-Puebla, C. and Challinor, A. J.: Effects of diurnal temperature range and drought on wheat yield in Spain, Theor. Appl. Climatol., 129(1–2), 503–519, doi:10.1007/s00704-016-1779-9, 2017.

Hernández-Barrera, S. and Rodríguez-Puebla, C.: Wheat yield in Spain and associated solar radiation patterns, Int. J. Climatol., 37(January), 45–58, doi:10.1002/joc.4975, 2017.

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B. and Stepanek, P.: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, Nat. Geosci., 4(1), 17–21, doi:10.1038/ngeo1032, 2011.

Hlavinka, P., Trnka, M., Eitzinger, J., Smutný, V., Thaler, S., Žalud, Z., Rischbeck, P. and Kr, J.: The performance of CERES-Barley and CERES-Wheat under various soil conditions and tillage practices in Central Europe, Bodenkultur, 61(1), 2010.

Van Hoolst, R., Eerens, H., Haesen, D., Royer, A., Bydekerke, L., Rojas, O., Li, Y. and Racionzer, P.: FAO's AVHRR-based Agricultural Stress Index System (ASIS) for global drought monitoring, Int. J. Remote Sens., 37(2), 418–439, doi:10.1080/01431161.2015.1126378, 2016.

Iglesias, A. and Quiroga, S.: Measuring the risk of climate variability to cereal production at five sites in Spain, Clim. Res., 34(1), 47–57, doi:10.3354/cr034047, 2007.

Iglesias, A., Garrote, L. and Martin-Carrasco, F.: Drought risk management in Mediterranean river basins., Integr. Environ. Assess. Manag., 5(1), 11–16, 2009.

Iglesias, A., Mougou, R., Moneo, M. and Quiroga, S.: Towards adaptation of agriculture to climate change in the Mediterranean, Reg. Environ. Chang., 11(SUPPL. 1), 159–166,

doi:10.1007/s10113-010-0187-4, 2011.

Iglesias, E., Báez, K. and Diaz-Ambrona, C. H.: Assessing drought risk in Mediterranean Dehesa grazing lands, Agric. Syst., 149, 65–74, doi:10.1016/j.agsy.2016.07.017, 2016.

IM and AEMET: Iberian climate atlas. Air Temperature and Precipitation (1971–2000), Inst. Meteorol. Port. Agencia Estatal Meteorol. Española, 2011.

Incerti, G., Feoli, E., Salvati, L., Brunetti, A. and Giovacchini, A.: Analysis of bioclimatic time series and their neural network-based classification to characterise drought risk patterns in South Italy, Int. J. Biometeorol., 51(4), 253–263, doi:10.1007/s00484-006-0071-6, 2007.

IPCC: Managing the risks of extreme events and disasters to advance climate change adaptation., 2012.

IPCC: Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (SRCCL), 2019.

Jayanthi, H., Husak, G. J., Funk, C., Magadzire, T., Adoum, A. and Verdin, J. P.: A probabilistic approach to assess agricultural drought risk to maize in Southern Africa and millet in Western Sahel using satellite estimated rainfall, Int. J. Disaster Risk Reduct., 10(PB), 490–502, doi:10.1016/j.ijdrr.2014.04.002, 2014.

Jiang, D., Yang, X., Clinton, N. and Wang, N.: An artificial neural network model for estimating crop yields using remotely sensed information, Int. J. Remote Sens., 25(9), 1723–1732, doi:10.1080/0143116031000150068, 2004.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge University Press., 2003.

El Kenawy, A., López-Moreno, J. I. and Vicente-Serrano, S. M.: Trend and variability of surface air temperature in northeastern Spain (1920-2006): Linkage to atmospheric circulation, Atmos. Res., 106, 159–180, doi:10.1016/j.atmosres.2011.12.006, 2012.

Kogan, F.: World Droughts in the New Millennium from AVHRR-based Vegetation Health Indices Principles of a New Algorithm, Eos EOS Trans., 83(48), 3–7, doi:10.1029/2002EO000382, 2002.

Kogan, F., Stark, R., Gitelson, a., Jargalsaikhan, L., Dugrajav, C. and Tsooj, S.: Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices, Int. J. Remote Sens., 25(14), 2889–2896, doi:10.1080/01431160410001697619, 2004.

Kogan, F., Yang, B., Guo, W., Pei, Z. and Jiao, X.: Modelling corn production in China using AVHRR-based vegetation health indices, Int. J. Remote Sens., 26(11), 2325–2336, doi:10.1080/01431160500034235, 2005.

Kogan, F., Guo, W., Strashnaia, A., Kleshenko, A., Chub, O. and Virchenko, O.: Modelling and prediction of crop losses from NOAA polar-orbiting operational satellites, Geomatics, Nat. Hazards Risk, 7(3), 886–900, doi:10.1080/19475705.2015.1009178, 2015a.

Kogan, F., Goldberg, M., Schott, T. and Guo, W.: Suomi NPP/VIIRS: improving drought watch, crop loss prediction, and food security, Int. J. Remote Sens., 1161(November), 1–11, doi:10.1080/01431161.2015.1095370, 2015b.

Kogan, F. N.: Remote sensing of weather impacts on vegetation in non-homogeneous areas, Int. J. Remote Sens., 11(8), 1405–1419, doi:10.1080/01431169008955102, 1990.

Kogan, F. N.: Application of vegetation index and brightness temperature for drought detection, Adv. Sp. Res., 15(11), 91–100, doi:10.1016/0273-1177(95)00079-T, 1995.

Kogan, F. N.: Global Drought Watch from Space, Bull. Am. Meteorol. Soc., 78(4), 621–636, doi:10.1175/1520-0477(1997)078<0621:GDWFS>2.0.CO;2, 1997.

Kogan, F. N.: Operational space technology for global vegetation assessment, Bull. Am. Meteorol. Soc., 82(9), 1949–1964, doi:10.1175/1520-0477(2001)082, 2001.

Lakshmi, V.: Remote Sensing of Hydrological Extremes. Springer Remote Sensing/Photogrammetry., 2016.

Le, J. A., El-Askary, H. M., Allali, M. and Struppa, D. C.: Application of recurrent neural networks for drought projections in California, Atmos. Res., doi:10.1016/j.atmosres.2017.01.002, 2017.

Lechenet, M., Makowski, D., Py, G. and Munier-Jolain, N.: Profiling farming management strategies with contrasting pesticide use in France, Agric. Syst., 149, 40–53, doi:10.1016/j.agsy.2016.08.005, 2016.

Lee, T., Modarres, R. and Ouarda, T. B. M. J.: Data-based analysis of bivariate copula tail dependence for drought duration and severity, Hydrol. Process., 27(10), 1454–1463, doi:10.1002/hyp.9233, 2013.

Leff, B., Ramankutty, N. and Foley, J. A.: Geographic distribution of major crops across the world, Global Biogeochem. Cycles, 18(1), n/a-n/a, doi:10.1029/2003GB002108, 2004.

Leng, G. and Hall, J.: Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future, Sci. Total Environ., 654, 811–821, doi:10.1016/j.scitotenv.2018.10.434, 2019.

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., Mcinnes, K., Risbey, J., Schuster, S., Jakob, D. and Stafford-Smith, M.: A compound event framework for understanding extreme impacts, Wiley Interdiscip. Rev. Clim. Chang., 5(1), 113–128, doi:10.1002/wcc.252, 2014.

Lesk, C., Rowhani, P. and Ramankutty, N.: Influence of extreme weather disasters on global crop production, Nature, 529(7584), 84–87, doi:10.1038/nature16467, 2016.

Li, Y., Ye, W., Wang, M. and Yan, X.: Climate Change and drought: a risk assessment of crop-yield impacts, Clim. Res., 39, 31–46, 2009.

Li, Y., Gu, W., Cui, W., Chang, Z. and Xu, Y.: Exploration of copula function use in crop meteorological drought risk analysis: a case study of winter wheat in Beijing, China, Nat. Hazards, 77(2), 1289–1303, doi:10.1007/s11069-015-1649-2, 2015.

Loboda, T., Krankina, O., Savin, I., Kurbanov, E. and Hall, J.: Land management and the impact of the 2010 extreme drought event on the agricultural and ecological systems of European Russia, in: Land-Cover and Land-Use Changes in Eastern Europe after the Collapse of the Soviet Union in 1991, Springer, 173–192, doi:10.1007/978-3-319-42638-9_8, 2017.

Lu, Y., Hu, H., Li, C. and Tian, F.: Increasing compound events of extreme hot and dry days during growing seasons of wheat and maize in China, Sci. Rep., 8(1), 1–8, doi:10.1038/s41598-018-34215-y, 2018.

Madadgar, S. and Moradkhani, H.: Drought Analysis under Climate Change Using Copula, J. Hydrol. Eng., 18(7), 746–759, doi:10.1061/(ASCE)HE.1943-5584 .0000532., 2013.

Madadgar, S., AghaKouchak, A., Farahmand, A. and Davis, S. J.: Probabilistic estimates of

drought impacts on agricultural production, Geophys. Res. Lett., 44(15), 7799–7807, doi:10.1002/2017GL073606, 2017.

Maity, R.: Statistical methods in hydrology, Springer Trans. Civ. Environ. Eng., doi:https://doi.org/10.1007/978-981-10-8779-0_10, 2018.

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A. F., Maraun, D. and Vrac, M.: Soil Moisture Drought in Europe: A Compound Event of Precipitation and Potential Evapotranspiration on Multiple Time Scales, J. Hydrometeorol., 19(8), 1255–1271, doi:10.1175/JHM-D-18-0017.1, 2018.

Martin-Vide, J. and Lopez-Bustins, J.: The Western Mediterranean Oscillation and rainfall in the Iberian Peninsula, Int. J. Climatol., 26, 1455–1475, doi:DOI: 10.1002/joc.1388 THE, 2006.

Martins, D. S., Raziei, T., Paulo, A. A. and Pereira, L. S.: Spatial and temporal variability of precipitation and drought in Portugal, Nat. Hazards Earth Syst. Sci., 12(5), 1493–1501, doi:10.5194/nhess-12-1493-2012, 2012.

Matsumura, K., Gaitan, C. F., Sugimoto, K., Cannon, A. J. and Hsieh, W. W.: Maize yield forecasting by linear regression and artificial neural networks in Jilin, China, J. Agric. Sci., 153(03), 399–410, doi:10.1017/S0021859614000392, 2015.

Mazdiyasni, O. and AghaKouchak, A.: Substantial increase in concurrent droughts and heatwaves in the United States, Proc. Natl. Acad. Sci., 112(37), 11484–11489, doi:10.1073/pnas.1422945112, 2015.

Mazdiyasni, O., AghaKouchak, A., Davis, S. J., Madadgar, S., Mehran, A., Ragno, E., Sadegh, M., Sengupta, A., Ghosh, S., Dhanya, C. T. and Niknejad, M.: Increasing probability of mortality during Indian heat waves, Sci. Adv., 3(6), 1–6, doi:10.1126/sciadv.1700066, 2017.

McKee, T. B., Doesken, N. J., Kleist, J. and others: The relationship of drought frequency and duration to time scales, in Proceedings of the 8th Conference on Applied Climatology, vol. 17, pp. 179–183., 1993.

Mirabbasi, R., Fakheri-Fard, A. and Dinpashoh, Y.: Bivariate drought frequency analysis using the copula method, Theor. Appl. Climatol., 108(1–2), 191–206, doi:10.1007/s00704-011-0524-7, 2012.

Miralles, D. G., Teuling, A. J., Van Heerwaarden, C. C. and De Arellano, J. V. G.: Mega-

heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation, Nat. Geosci., 7(5), 345–349, doi:10.1038/ngeo2141, 2014.

Miralles, D. G., Gentine, P., Seneviratne, S. I. and Teuling, A. J.: Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges, Ann. N. Y. Acad. Sci., 1436(1), 19–35, doi:10.1111/nyas.13912, 2019.

Mishra, A. K., Ines, A. V. M., Das, N. N., Prakash Khedun, C., Singh, V. P., Sivakumar, B. and Hansen, J. W.: Anatomy of a local-scale drought: Application of assimilated remote sensing products, crop model, and statistical methods to an agricultural drought study, J. Hydrol., 526, 15–29, doi:10.1016/j.jhydrol.2014.10.038, 2015.

Moreira, E., Pires, C. and Moreira, E.: SPI drought class predictions driven by NAO index using log- linear modeling, Water, 7, doi:10.3390/w70x000x, 2015.

Moreira, E. E., Mexia, J. T. and Pereira, L. S.: Assessing homogeneous regions relative to drought class transitions using an ANOVA-like inference. Application to Alentejo, Portugal, Stoch. Environ. Res. Risk Assess., 27(1), 183–193, doi:10.1007/s00477-012-0575-z, 2013.

Morid, S., Smakhtin, V. and Bagherzadeh, K.: Drought forecasting using artificial neural networks and time series of drought indices, Int. J. Climatol., 27(15), 2103–2111, doi:10.1002/joc.1498, 2007.

Mueller, B. and Seneviratne, S. I.: Hot days induced by precipitation deficits at the global scale, Proc. Natl. Acad. Sci. U. S. A., 109(31), 12398–12403, doi:10.1073/pnas.1204330109, 2012.

Muñoz-Díaz, D. and Rodrigo, F. S.: Seasonal rainfall variations in Spain (1912-2000) and their links to atmospheric circulation, Atmos. Res., 81(1), 94–110, doi:10.1016/j.atmosres.2005.11.005, 2006.

Murthy, C. S., Laxman, B. and Sesha Sai, M. V. R.: Geospatial analysis of agricultural drought vulnerability using a composite index based on exposure, sensitivity and adaptive capacity, Int. J. Disaster Risk Reduct., 12, 163–171, doi:10.1016/j.ijdrr.2015.01.004, 2015.

NASA Earth Observatory: Drought Exacerbates Australian Fires, https://earthobservatory.nasa.gov/images/145599/drought-exacerbates-australian-fires, 2019a.

NASA Earth Observatory: Fires in Greece, https://earthobservatory.nasa.gov/images/39913/fires-in-greece, 2019b.

Naumann, G., Spinoni, J., Vogt, J. and Barbosa, P.: Assessment of drought impacts and their uncertainties in Europe, Environ. Res. Lett., 10, 124013, doi:10.1088/1748-9326/10/12/124013, 2015.

Nelsen, R. B.: An Introduction to Copulas, Springer Ser. Stat., 53(9), 276, doi:10.1017/CBO9781107415324.004, 2006.

Nguyen, T., Mula, L., Cortignani, R., Seddaiu, G., Dono, G., Virdis, S., Pasqui, M. and Roggero, P.: Perceptions of Present and Future Climate Change Impacts on Water Availability for Agricultural Systems in the Western Mediterranean Region, Water, 8(11), 523, doi:10.3390/w8110523, 2016.

Ojeda, M. G.-V., Jiménez, E. R., Gámiz-Fortis, S. R., Castro-Díez, Y. and Parra, M. J. E.: Understanding the Drought Phenomenon in the Iberian Peninsula, Intech, i(tourism), 13, doi:http://dx.doi.org/10.5772/57353, 2016.

Okhrin, O. and Ristig, A.: Hierarchical archimedean copulae: The HAC package, J. Stat. Softw., 58(4), 1–20, doi:10.18637/jss.v058.i04, 2014.

Palmer, W. C.: Meteorological Drought, U.S. Weather Bur. Res. Pap. No. 45, 58 [online] Available from: https://www.ncdc.noaa.gov/temp-and-precip/drought/docs/palmer.pdf, 1965.

Paredes, P., Rodrigues, G. C., Alves, I. and Pereira, L. S.: Partitioning evapotranspiration, yield prediction and economic returns of maize under various irrigation management strategies, Agric. Water Manag., 135, 27–39, doi:10.1016/j.agwat.2013.12.010, 2014.

Paredes, P., Rodrigues, G. C., Cameira, M. do R., Torres, M. O. and Pereira, L. S.: Assessing yield, water productivity and farm economic returns of malt barley as influenced by the sowing dates and supplemental irrigation, Agric. Water Manag., doi:10.1016/j.agwat.2016.05.033, 2016.

Páscoa, P., Gouveia, C. M., Russo, A. and Trigo, R. M.: Drought trends in the Iberian Peninsula over the last 112 years, Adv. Meteorol., 2017, doi:10.1155/2017/4653126, 2017a.

Páscoa, P., Gouveia, C. M., Russo, A. and Trigo, R. M.: The role of drought on wheat yield interannual variability in the Iberian Peninsula from 1929 to 2012, Int. J. Biometeorol., 61(3), 439–451, doi:10.1007/s00484-016-1224-x, 2017b.

Peña-Gallardo, M., Vicente-Serrano, S. M., Domínguez-Castro, F. and Beguería, S.: The

impact of drought on the productivity of two rainfed crops in Spain, Nat. Hazards Earth Syst. Sci. Discuss., 1–30, doi:10.5194/nhess-2019-1, 2019.

Péres, W. E., Ribeiro, A. F. S., Russo, A. and Nunes, B.: The association between air temperature and mortality in two brazilian health regions, Climate, 8(2), 1–14, doi:10.3390/cli8020032, 2020.

Pires, C. A. L. and Ribeiro, A. F. S.: Separation of the atmospheric variability into non-Gaussian multidimensional sources by projection pursuit techniques, Clim. Dyn., 1–30, doi:10.1007/s00382-016-3112-9, 2016.

Popova, Z., Ivanova, M., Martins, D., Pereira, L. S., Doneva, K., Alexandrov, V. and Kercheva, M.: Vulnerability of Bulgarian agriculture to drought and climate variability with focus on rainfed maize systems, Nat. Hazards, 74(2), 865–886, doi:10.1007/s11069-014-1215-3, 2014.

Quesada, B., Vautard, R., Yiou, P., Hirschi, M. and Seneviratne, S. I.: Asymmetric European summer heat predictability from wet and dry southern winters and springs, Nat. Clim. Chang., 2(10), 736–741, doi:10.1038/nclimate1536, 2012.

Quiroga, S. and Iglesias, A.: A comparison of the climate risks of cereal, citrus, grapevine and olive production in Spain, Agric. Syst., 101(1–2), 91–100, doi:10.1016/j.agsy.2009.03.006, 2009.

Quiroga, S., Fernández-Haddad, Z. and Iglesias, A.: Crop yields response to water pressures in the Ebro basin in Spain: Risk and water policy implications, Hydrol. Earth Syst. Sci., 15(2), 505–518, doi:10.5194/hess-15-505-2011, 2011.

Reidsma, P., Ewert, F., Lansink, A. O. and Leemans, R.: Adaptation to climate change and climate variability in European agriculture: The importance of farm level responses, Eur. J. Agron., 32(1), 91–102, doi:10.1016/j.eja.2009.06.003, 2010.

Resco, P., Quiroga, S., Iglesias, A. and Sotés, V.: Risk of Climate Change for Grapevine Production in Mediterranean Areas, Le Bull. l'OIV, 83(953-954–955), 315, 2010.

Ribeiro, A. F. S., Russo, A., Gouveia, C. M. and Páscoa, P.: Copula-based agricultural drought risk of rainfed cropping systems, Agric. Water Manag., 223, doi:10.1016/j.agwat.2019.105689, 2019a.

Ribeiro, A. F. S., Russo, A., Gouveia, C. M. and Páscoa, P.: Modelling drought-related yield

losses in Iberia using remote sensing and multiscalar indices, Theor. Appl. Climatol., 136, 203–220, doi:10.1007/s00704-018-2478-5, 2019b.

Ribeiro, A. F. S., Russo, A., Gouveia, C. M., Páscoa, P. and Pires, C. A. L.: Probabilistic modelling of the dependence between rainfed crops and drought hazard, Nat. Hazards Earth Syst. Sci., 19(12), 2795–2809, doi:10.5194/nhess-19-2795-2019, 2019c.

Ribeiro, A. F. S., Russo, A., Gouveia, C. M., Páscoa, P. and Zscheischler, J.: Risk of crop failure due to compound dry and hot extremes estimated with nested copulas, Biogeosciences Discuss., 1–20, doi:https://doi.org/10.5194/bg-2020-116, in review, preprint, 2020.

Ribeiro, A. F. S., Russo, A., Gouveia, C. M. and Pires, C. A. L.: Drought-related hot summers: a joint probability analysis in the Iberian Peninsula, Weather Clim. Extrem., in review, n.d.

Rodriguez-Puebla, C., Encinas, a H., Nieto, S. and Garmendia, J.: Spatial and temporal patterns of annual precipitation variability over the Iberian Peninsula, Int. J. Climatol., 18(3), 299–316, doi:10.1002/(SICI)1097-0088(19980315)18:3<299::AID-JOC247>3.0.CO;2-L, 1998.

Rodríguez-Puebla, C., Encinas, A. H., García-Casado, L. A. and Nieto, S.: Trends in warm days and cold nights over the Iberian Peninsula: Relationships to large-scale variables, Clim. Change, 100(3), 667–684, doi:10.1007/s10584-009-9721-0, 2010.

Rodríguez Díaz, J. A., Weatherhead, E. K., Knox, J. W. and Camacho, E.: Climate change impacts on irrigation water requirements in the Guadalquivir river basin in Spain, Reg. Environ. Chang., 7(3), 149–159, doi:10.1007/s10113-007-0035-3, 2007.

Rojas, O., Vrieling, A. and Rembold, F.: Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery, Remote Sens. Environ., 115(2), 343–352, doi:10.1016/j.rse.2010.09.006, 2011.

Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E. and Bloomfield, J.: Increased crop damage in the US from excess precipitation under climate change, Glob. Environ. Chang., 12(3), 197–202, doi:10.1016/S0959-3780(02)00008-0, 2002.

Ruiz-Ramos, M. and Mínguez, M. I.: Evaluating uncertainty in climate change impacts on crop productivity in the Iberian Peninsula, Clim. Res., 44(1), 69–82, doi:10.3354/cr00933, 2010.

Russo, A., Raischel, F. and Lind, P. G.: Air quality prediction using optimal neural networks with stochastic variables, Atmos. Environ., 79, 822–830, doi:10.1016/j.atmosenv.2013.07.072,

2013.

Russo, A., Lind, P. G., Raischel, F., Trigo, R. and Mendes, M.: Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales, Atmos. Pollut. Res., 6(3), 540–549, doi:10.5094/APR.2015.060, 2015a.

Russo, A., Gouveia, C. M., Páscoa, P., DaCamara, C. C., Sousa, P. M. and Trigo, R. M.: Assessing the role of drought events on wildfires in the Iberian Peninsula, Agric. For. Meteorol., 237–238, doi:10.1016/j.agrformet.2017.01.021, 2017.

Russo, A., Gouveia, C. M., Dutra, E., Soares, P. M. M. and Trigo, R. M.: The synergy between drought and extremely hot summers in the Mediterranean, Environ. Res. Lett., 14(1), 014011, doi:10.1088/1748-9326/aaf09e, 2019.

Russo, A. C., Gouveia, C. M., Trigo, R. M., Liberato, M. L. R. and DaCamara, C. C.: The influence of circulation weather patterns at different spatial scales on drought variability in the Iberian Peninsula, Front. Environ. Sci., 3(February), 1+, doi:10.3389/fenvs.2015.00001, 2015b.

Russo, S., Sillmann, J. and Fischer, E. M.: Top ten European heatwaves since 1950 and their occurrence in the coming decades, Environ. Res. Lett., 10(12), doi:10.1088/1748-9326/10/12/124003, 2015c.

Saadi, S., Todorovic, M., Tanasijevic, L., Pereira, L. S., Pizzigalli, C. and Lionello, P.: Climate change and Mediterranean agriculture: Impacts on winter wheat and tomato crop evapotranspiration, irrigation requirements and yield, Agric. Water Manag., 147, 103–115, doi:10.1016/j.agwat.2014.05.008, 2015.

Saghafian, B. and Mehdikhani, H.: Drought characterization using a new copula-based trivariate approach, Nat. Hazards, 72(3), 1391–1407, doi:10.1007/s11069-013-0921-6, 2014.

Salvadori, G. and De Michele, C.: On the Use of Copulas in Hydrology: Theory and Practice, J. Hydrol. Eng., 12(4), 369–380, doi:10.1061/(ASCE)1084-0699(2007)12:4(369), 2007.

Schumacher, D. L., Keune, J., van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Teuling, A. J. and Miralles, D. G.: Amplification of mega-heatwaves through heat torrents fuelled by upwind drought, Nat. Geosci., doi:10.1038/s41561-019-0431-6, 2019.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B. and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A

review, Earth-Science Rev., 99(3–4), 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010.

Serinaldi, F.: An uncertain journey around the tails of multivariate hydrological distributions, Water Resour. Res., 49(10), 6527–6547, doi:10.1002/wrcr.20531, 2013.

Serinaldi, F.: Can we tell more than we can know? The limits of bivariate drought analyses in the United States, Stoch. Environ. Res. Risk Assess., 30(6), 1691–1704, doi:10.1007/s00477-015-1124-3, 2016.

Serinaldi, F., Bonaccorso, B., Cancelliere, A. and Grimaldi, S.: Probabilistic characterization of drought properties through copulas, Phys. Chem. Earth, 34(10–12), 596–605, doi:10.1016/j.pce.2008.09.004, 2009.

Skakun, S., Kussul, N., Shelestov, A. and Kussul, O.: The use of satellite data for agriculture drought risk quantification in Ukraine, Geomatics, Nat. Hazards Risk, 7(3), 901–917, doi:10.1080/19475705.2015.1016555, 2016.

Sklar, A.: Fonctions de Répartition à n Dimensions et Leurs Marges, Inst. Stat. l'Université Paris, 8, 229–231, 1959.

Sousa, P. M., Trigo, R. M., Aizpurua, P., Nieto, R., Gimeno, L. and Garcia-Herrera, R.: Trends and extremes of drought indices throughout the 20th century in the Mediterranean, Nat. Hazards Earth Syst. Sci., 11(1), 33–51, doi:10.5194/nhess-11-33-2011, 2011.

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M. M. and Santos, J. A.: European temperature responses to blocking and ridge regional patterns, Clim. Dyn., 50(1–2), 457–477, doi:10.1007/s00382-017-3620-2, 2018.

Sousa, P. M., Barriopedro, D., Ramos, A. M., García-herrera, R., Espírito-santo, F. and Trigo, R. M.: Saharan air intrusions as a relevant mechanism for Iberian heatwaves: The record breaking events of August 2018 and June 2019, Weather Clim. Extrem., (June), 100224, doi:10.1016/j.wace.2019.100224, 2019.

Tomczyk, A. M., Pólrołniczak, M. and Bednorz, E.: Circulation conditions' effect on the occurrence of heatwaves in Western and Southwestern Europe, Atmosphere (Basel)., 8(2), doi:10.3390/atmos8020031, 2017.

Trigo, R. M., Pereira, J. M. C., Pereira, M. G., Mota, B., Calado, T. J., DaCamara, C. C. and Santo, F. E.: Atmospheric conditions associated with the exceptional fire season of 2003 in

Portugal, Int. J. Climatol., 26, 1741–1757, doi:10.1002/joc, 2006.

Trigo, R. M., Añel, J. A., Barriopedro, D., García-Herrera, R., Gimeno, L., Nieto, R., Castillo, R., Allen, M. R. and Massey, N.: The Record Winter Drought of 2011 – 12 in the Iberian Peninsula, Am. Meteorol. Soc., (September), 41–45, 2013.

Vergni, L., Todisco, F. and Mannocchi, F.: Analysis of agricultural drought characteristics through a two-dimensional copula, Water Resour. Manag., 29(8), 2819–2835, doi:10.1007/s11269-015-0972-4, 2015.

Vicente-Serrano, S. M.: Differences in spatial patterns of drought on different time scales: An analysis of the Iberian Peninsula, Water Resour. Manag., 20(1), 37–60, doi:10.1007/s11269-006-2974-8, 2006.

Vicente-Serrano, S. M.: Evaluating the impact of drought using remote sensing in a Mediterranean, Semi-arid Region, Nat. Hazards, 40(1), 173–208, doi:10.1007/s11069-006-0009-7, 2007.

Vicente-Serrano, S. M. and Cuadrat-Prats, J. M.: Trends in drought intensity and variability in the middle Ebro valley (NE of the Iberian peninsula) during the second half of the twentieth century, Theor. Appl. Climatol., 88(3–4), 247–258, doi:10.1007/s00704-006-0236-6, 2007.

Vicente-Serrano, S. M., Cuadrat-Prats, J. M. and Romo, A.: Early prediction of crop production using drought indices at different timescales and remote sensing data: application in the Ebro Valley (north-east Spain), Int. J. Remote Sens., 27(3), 511–518, doi:10.1080/01431160500296032, 2006.

Vicente-Serrano, S. M., Begueria, S. and Lopez-Moreno, J. I.: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index, J. Clim., 23(7), 1696–1718, doi:10.1175/2009JCLI2909.1, 2010.

Vicente-Serrano, S. M., López-Moreno, J. I., Drumond, A., Gimeno, L., Nieto, R., Morán-Tejeda, E., Lorenzo-Lacruz, J., Beguería, S. and Zabalza, J.: Effects of warming processes on droughts and water resources in the NW Iberian Peninsula (1930-2006), Clim. Res., 48(2–3), 203–212, doi:10.3354/cr01002, 2011.

Vicente-Serrano, S. M., Gouveia, C., Camarero, J. J., Beguería, S., Trigo, R., López-Moreno, J. I., Azorín-Molina, C., Pasho, E., Lorenzo-Lacruz, J., Revuelto, J. and others: Response of

vegetation to drought time-scales across global land biomes, Proc. Natl. Acad. Sci., 110(1), 52–57, 2013.

Vicente-Serrano, S. M., Lopez-Moreno, J.-I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., Azorin-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F. and Espejo, F.: Evidence of increasing drought severity caused by temperature rise in southern Europe, Environ. Res. Lett., 9(4), 044001, doi:10.1088/1748-9326/9/4/044001, 2014.

Vogel, M. M., Zscheischler, J. and Seneviratne, S. I.: Varying soil moisture-atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central Europe, Earth Syst. Dyn., 9(3), 1107–1125, doi:10.5194/esd-9-1107-2018, 2018.

Warner and Thomas, T.: Numerical Weather and Climate Prediction Cambridge, Cambridge University Press., 2011.

Wells, N., Goddard, S. and Hayes, M. J.: A self-calibrating Palmer Drought Severity Index, J. Clim., 17(12), 2335–2351, doi:10.1175/1520-0442(2004)017<2335:ASPDSI>2.0.CO;2, 2004.

Whan, K., Zscheischler, J., Orth, R., Shongwe, M., Rahimi, M., Asare, E. O. and Seneviratne, S. I.: Impact of soil moisture on extreme maximum temperatures in Europe, Weather Clim. Extrem., 9, doi:10.1016/j.wace.2015.05.001, 2015.

Wilhelmi, O. V and Wilhite, D. A.: Assessing Vulnerability to Agricultural Drought : A Nebraska Case Study, , 25(13029), 37–58, 2002.

Wilhite, D.: Drought, Encycl. World Climatol., 338–341, doi:10.1007/1-4020-3266-8_70, 2005.

Wilhite, D. A.: Drought as a natural hazard: Concepts and definitions, Drought A Glob. Assess., 3–18, 2000.

Wilhite, D. A. and Glantz, M. H.: Understanding the Drought Phenomena: The Role of Definitions*, Water Int., 10(3), 111–120, 1985.

Wilks, D.: Statistical Methods in the Atmospheric Sciences, 2nd Ed., Academic Press, Oxford, UK., 2006.

Wu, H. and Wilhite, D. A.: An Operational Agricultural Drought Risk Assessment Model for

Nebraska , USA, Nat. Hazards, 33, 1–21, 2004.

Wu, X., Hao, Z., Hao, F., Li, C. and Zhang, X.: Spatial and Temporal Variations of Compound Droughts and Hot Extremes in China, Atmosphere (Basel)., 10(2), 95, doi:10.3390/atmos10020095, 2019.

Xie, W., Xiong, W., Pan, J., Ali, T., Cui, Q., Guan, D., Meng, J., Mueller, N. D., Lin, E. and Davis, S. J.: Decreases in global beer supply due to extreme drought and heat, Nat. Plants, doi:10.1038/s41477-018-0263-1, 2018.

Yang, C., Fraga, H., van Ieperen, W., Trindade, H. and Santos, J. A.: Effects of climate change and adaptation options on winter wheat yield under rainfed Mediterranean conditions in southern Portugal, Clim. Change, 154(1–2), 159–178, doi:10.1007/s10584-019-02419-4, 2019.

Yuan, W., Cai, W., Chen, Y., Liu, S., Dong, W., Zhang, H., Yu, G., Chen, Z., He, H., Guo, W., Liu, D., Liu, S., Xiang, W., Xie, Z., Zhao, Z. and Zhou, G.: Severe summer heatwave and drought strongly reduced carbon uptake in Southern China, Sci. Rep., 6(January 2015), 1–12, doi:10.1038/srep18813, 2016.

Zampieri, M., D'Andrea, F., Vautard, R., Ciais, P., De Noblet-Ducoudré, N. and Yiou, P.: Hot European summers and the role of soil moisture in the propagation of mediterranean drought, J. Clim., 22(18), 4747–4758, doi:10.1175/2009JCLI2568.1, 2009.

Zampieri, M., Ceglar, A., Dentener, F. and Toreti, A.: Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales, Environ. Res. Lett., 12(6), 064008, doi:10.1088/1748-9326/aa723b, 2017.

Zarei, R., Sarajian, M. and Bazgeer, S.: Monitoring Meteorological Drought in Iran Using Remote Sensing and Drought Indices, DESERT, 18, 89–97, 2013.

Zargar, A., Sadiq, R., Naser, B. and Khan, F. I.: A review of drought indices, Environ. Rev., 19(NA), 333–349, doi:10.1139/a11-013, 2011.

Zhang, Q., Chen, Y. D., Chen, X. and Li, J.: Copula-Based Analysis of Hydrological Extremes and Implications of Hydrological Behaviors in the Pearl River Basin, China, J. Hydrol. Eng., 16(7), 598–607, doi:10.1061/(asce)he.1943-5584.0000350, 2011a.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B. and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and

precipitation data, Wiley Interdiscip. Rev. Clim. Chang., 2(6), 851–870, doi:10.1002/wcc.147, 2011b.

Zscheischler, J. and Seneviratne, S. I.: Dependence of drivers affects risks associated with compound events, Sci. Adv., 3(6), 1–11, doi:10.1126/sciadv.1700263, 2017.

Zscheischler, J., Michalak, A. M., Schwalm, C., Mahecha, M. D., Huntzinger, D. N., Reichstein, M., Berthier, G., Ciais, P., Cook, R. B., El-Masri, B., Huang, M., Ito, A., Jain, A., King, A., Lei, H., Lu, C., Mao, J., Peng, S., Poulter, B., Ricciuto, D., Shi, X., Tao, B., Tian, H., Viovy, N., Wang, W., Wei, Y., Yang, J. and Zeng, N.: Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data, Global Biogeochem. Cycles, 28(6), 585–600, doi:10.1002/2014GB004826, 2014.

Zscheischler, J., Orth, R. and Seneviratne, S. I.: Bivariate return periods of temperature and precipitation explain a large fraction of European crop yields, Biogeosciences, 14(13), 3309–3320, doi:10.5194/bg-14-3309-2017, 2017.

Zscheischler, J., Westra, S., Van Den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., Aghakouchak, A., Bresch, D. N., Leonard, M., Wahl, T. and Zhang, X.: Future climate risk from compound events, Nat. Clim. Chang., 8(6), 469–477, doi:10.1038/s41558-018-0156-3, 2018.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., R., C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N., Thiery, W. and Vignotto, E.: A typology of compound weather and climate events, Nat. Rev. Earth Environ., 2020.